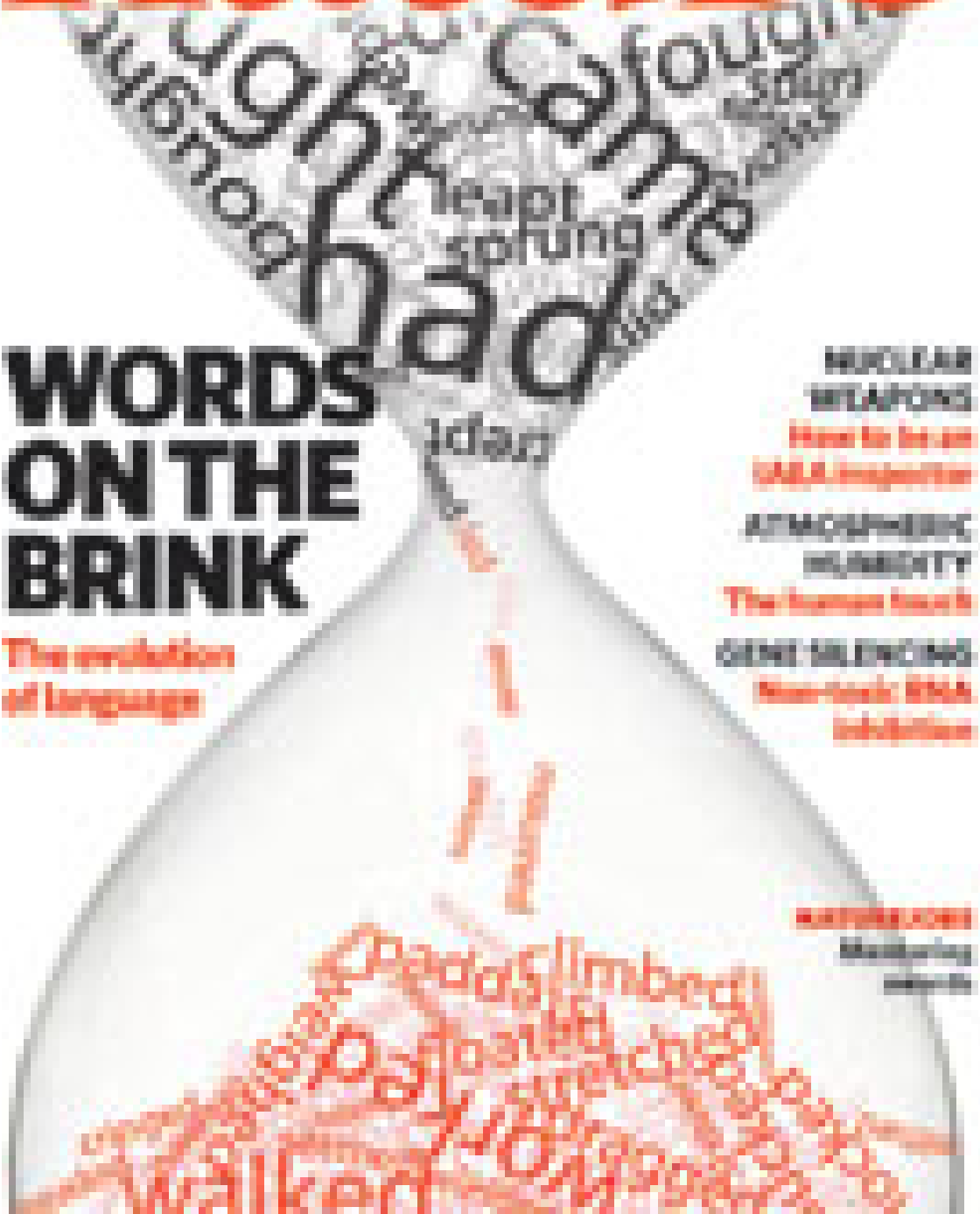
nature



WORDS ON THE BRINK

The evolution of language

NUCLEAR WEAPONS
How to be an IAEA inspector

ATMOSPHERIC HUMIDITY
The human touch

GENE SILENCING
Non-toxic RNA inhibition

nature

www.nature.com/nature

Vol 449 | Issue no. 7163 | 11 October 2007

Kill king corn

Biofuels need new technology, new agronomy and new politics if they are not to do more harm than good.

Zea mays has become the very emblem of plenty, with rich golden cobs of corn (maize) overspilling from some of the most effectively farmed arable lands on the planet. *Jatropha curcas*, on the other hand, is an unprepossessing and indeed toxic plant, better suited to scrubland and hedges. Yet in the world of biofuels, ugly-duckling jatropha has the potential to be, if not a hero, then at least one of the good guys, and a harbinger of better things to come. The golden-headed siren corn, on the other hand, is inspiring a wrong-headed gold-rush — to a dead-end of development that is polluting the modest aspirations the world might have for biofuels in general.

The common complaints about biofuels — and they seem to become more common by the day — are that they are expensive and ineffective at reducing fossil-fuel consumption, that they intensify farming needlessly, that they dress up discredited farm subsidies in new green clothes, and that they push up the price of food. All these things are true to some extent of corn-based ethanol, America's biofuel of choice, and many are also true of Europe's favoured biodiesel plans.

As far as the greenhouse goes, figures from the International Institute for Sustainable Development's Global Subsidies Initiative put the cost of averting carbon dioxide emissions by using corn-based ethanol at more than $500 a tonne of carbon dioxide. What's more, the heavy use of nitrogen fertilizer in growing corn leads to significant emissions of nitrous oxide, an even more potent greenhouse gas.

Despite this, the generous tax allowance of 51 cents a gallon given to ethanol blenders in the United States has made corn peculiarly profitable (provided that tariffs continue to keep out far more efficiently produced ethanol from the sugar plantations of Brazil). In a recent article in *Foreign Affairs*, C. Ford Runge and Benjamin Senauer of the University of Minnesota in Minneapolis point to estimates that this artificial price-hike will drive world corn prices up by 20% by 2010. This has a knock-on effect on other staple crops — more land for corn means less for wheat, for example. Higher prices are good news for farmers, including some of those in developed countries. But they can be bad news for the very poor, who spend a disproportionate amount of their income on food. According to World Bank studies, for the poorest people in the world a 1% increase in the price of staple food leads to a 0.5% drop in caloric consumption.

This sorry state of affairs has the small benefit of providing a stark, contrasting background against which to sketch out what a successful and sustainable biofuels industry might look like. It will be based not on digestible starch from staple crops such as corn or cassava, but for the most part on indigestible cellulose, with some room for biodiesels that, because they grow on marginal land, do not compete with food production. In the medium to long term, it will not seek to produce ethanol — a poor fuel — but a range of more complex fuels delivered by carefully designed microbes.

"A successful biofuels industry will not be based on digestible starch from staple crops such as corn."

A rosy biofuels future will enjoy the benefits of free trade, allowing the countries and peoples of the tropics to ship some of their abundant sunlight north in the form of fuel. It will also require serious amounts of agronomic research — as we report on page 652, one of the most significant problems with jatropha is that, as yet, remarkably little is known about how best to grow and improve it. One focus of such research must be in the development of plants, such as jatropha, that make do on little water, and those that require low inputs of nitrogen. This is inherently more feasible in the case of fuels, where all that needs to be taken out of the system are carbon and hydrogen, than in the case of food, where there is a need to export nitrogen in the form of protein as well. Another focus will be on systems that actively store carbon in the soil, improving it for future agricultural use and at the same time doing a little bit more to take the edge off the carbon/climate crisis.

Biofuels are unlikely ever to be more than bit-players in the great task of weaning civilization from Earth's coal-mine and oil-well teats. But they may yet have valuable niches — including some that allow them to serve some of the world's poor, both as fuels for their own use and as exports. Provided, that is, that someone kills king corn. ■

A matter of trust

Social scientists studying electronic interactions must take the lead on preserving data security.

For a certain sort of social scientist, the traffic patterns of millions of e-mails look like manna from heaven. Such data sets allow them to map formal and informal networks and pecking orders, to see how interactions affect an organization's function, and to watch these elements evolve over time. They are emblematic of the vast amounts of structured information opening up new ways to study communities and societies. Such research could provide much-needed insight into some of the most pressing issues of our day, from the functioning of religious fundamentalism to the way behaviour influences epidemics.

One factor such research could helpfully focus on is the generation and transmission of trust. From the promise on a banknote to the exchange of rings at a wedding, our societies are based on the creation and protection of trust. More parochially, trust is of crucial importance to the contract between scientific expertise and the broader society that supports it. When it breaks down — whether over vaccines, nuclear waste or the security of the food chain — there are serious repercussions on both sides. The

analysis of large-scale social interactions in a way that reveals something about how trust functions in them is a fascinating direction for research.

But for such research to flourish, it must engender that which it seeks to describe. And so it is encouraging that computational social scientists are trying to anticipate threats to trust that are implicit in their work. Any data on human subjects inevitably raise privacy issues (see page 644), and the real risks of abuse of such data are difficult to quantify. But although the risks posed by researchers seem far lower than those posed by governments, private companies and criminals, their soul-searching is justified. Abuse or sloppiness could do untold damage to the emerging field.

Rules are needed to ensure data can be safely and routinely shared among scientists, thus avoiding a Wild West where researchers compete for key data sets no matter what the terms. The complexities of anonymizing data correctly, and the lack of experience of local ethical committees in such matters, calls for an institutionalized approach to setting standards and protocols for using personal data, as rightly recommended recently by the US National Academy of Sciences. Solid and well thought out rules for research are essential for building trust.

But researchers are right to argue that better protection ultimately lies in protecting electronic privacy more broadly. At present, privacy legislation in most countries lags far behind what is actually happening online. Companies and governments are insufficiently liable for abuse of the data they collect. Deliberate abuse of phone, e-mail, financial, medical and other personal records held should be made a criminal offence. Only by restoring a sense of control and ownership to the data subjects can better electronic trust be established. And only then will people come to feel more comfortable with the idea of research being done on their intimate details of their anonymized search results, e-mails, movements and telephone calls. ■

"Trust is of crucial importance to the contract between scientific expertise and the broader society that supports it."

How to be good?

Mentoring and training for ethical behaviour aren't all they're cracked up to be.

Researchers have always depended on their seniors to convey the peculiar knowledge of the lab. Techniques, values, scientific judgement and survival skills are imparted by good mentors at the bench and through challenging discourse at lab meetings or in the local pub. Young scientists enjoying such inductions are popularly viewed as the lucky ones, as opposed to those reared on the 'sink or swim' principle.

Some particularly grateful protégés offered nominations for *Nature*'s annual competition to find outstanding mentors, this year held in South Africa, and we are happy to congratulate the winners (see page 752). But the glow of a job well done should not stop us from asking questions about mentoring in general. Just how effective is it? In particular, do young scientists behave better as a result?

Melissa Anderson and her colleagues at the University of Minnesota in Minneapolis decided to investigate the relative effects of mentoring and formal instruction in setting a young scientist's ethical framework. They analysed the 3,250 respondents from a 2002 survey of about 6,900 grantees of the US National Institutes of Health (NIH) who were asked about their formal ethics instruction and the informal mentoring they had received — and how these had affected their subsequent behaviour (M. S. Anderson *et al. Acad. Med.* **82,** 853–860; 2007).

It turns out that, despite the NIH requiring formal training in the responsible conduct of research since 1990, it was still lacking ten years later. As many as a quarter of the NIH PhD fellows in 2000–01 had not taken ethics courses or been mentored in ethics workshops or discussion roundtables. One quarter of the survey respondents admitted that they did not feel well prepared to deal with ethical issues in their work.

More positively, about 90% of both early- and late-career respondents had discussed ethics with their mentors or colleagues and had been mentored in good research practices. Among younger scientists, biologists were among the least likely to have been mentored in ethics. It seems they received more mentoring in getting financial support for their work. Social and physical science postdocs were more likely to be mentored in how to survive in their fields and develop professional relationships.

So does the extent of mentorship and/or formal ethics training correlate with behaviour? The survey asked participants to report various types of problematic behaviours. Formal instruction exerted a disappointing influence on the early-career scientists: in fact, it significantly increased the odds of poor choices when collecting and analysing data, dealing with other researchers' confidential information or allowing funders inappropriate influence. Formal training was also correlated with a higher likelihood of not giving proper credit to others. How could this be? Perhaps such courses introduce scenarios that were unimaginable beforehand while suggesting that others have got away with such behaviours.

The results of mentoring early-career scientists were better but still mixed. Research mentoring (teaching good practice and presentation of one's results) and ethics mentoring decreased the likelihood of bad behaviour in almost all categories. But receiving mentoring advice on how to survive in the field and form professional relationships, or on how to support one's research, increased misbehaviour.

Collective discussions are perhaps a better way to reinforce good behaviour. A good forum for such discussions might be the lab's journal club — particularly if the mentor is skilled in inviting open discussion in a non-confrontational atmosphere. Mentors who regularly set aside time to discuss issues of concern may find that everyone benefits from the 'wisdom of the crowd'. Students can be asked to propose topics to kick off discussion: "I have a friend whose adviser asked her to write several anonymous reviews for him because he is too busy. Should she say yes? Ask for proper attribution?" Figuring it out together is the way to raise everyone's game. ■

RESEARCH HIGHLIGHTS

Tidal stream

Astrophys. J. **668,** 245–267 (2007)

Massive galaxies are thought to have formed through the merger of many smaller galaxies. In a new study, Karoline Gilbert at the University of California, Santa Cruz, and her colleagues identify a scar of such a process in the Andromeda galaxy (pictured), the nearest large galaxy to the Milky Way.

The researchers' survey of about 1,000 stars in Andromeda revealed a slow-moving stream of stars in the southeastern fringes of the galaxy. It is likely that this substructure is linked to the previously identified 'giant southern stream'. Computer simulations that explain the giant stream as the result of a galactic collision had predicted that such a feature should exist as an extended thread of debris.

R. GENDLER

CELL BIOLOGY

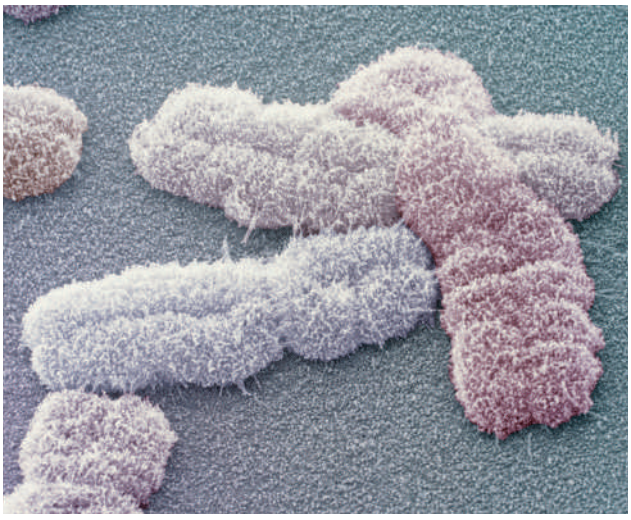
More than a safety cap

Science doi:10.1126/science.1147182 (2007)

Telomeres, the DNA 'caps' that protect the ends of chromosomes (pictured below), produce RNA, according to Joachim Lingner of the Swiss Federal Institute of Technology in Lausanne and his colleagues.

The discovery comes as a surprise. Telomeres are believed to have a role in ageing — they shorten as cells grow old, and cancer cells achieve immortality by maintaining long telomeres — but their highly repetitive DNA sequences were not thought to yield RNA.

Lingner and his team report that some of the telomere RNA remains clustered around telomeres. This trend was enhanced when researchers lowered the expression of several proteins involved in stabilizing telomeres, indirectly suggesting that the RNA may be involved in regulation of telomere length. The RNA's precise function, however, remains unclear.



A. SYRED/SPL

BIOCHEMISTRY

Second signal

Nature Chem. Biol. doi:10.1038/nchembio.2007.33 (2007)

Researchers in Japan have found a previously unknown player in the pathway by which nitric oxide helps cells to communicate at a molecular level.

Nitric oxide is a poisonous gas, but in minuscule quantities it has a multitude of roles in biochemical processes ranging from neuronal and vascular signalling to immunity and the regulation of cell death. The nitric oxide pathway is targeted by some drugs meant to regulate blood flow, including Viagra.

Takaaki Akaike at Kumamoto University and his colleagues describe how a nucleic acid, guanosine 3′,5′-cyclic monophosphate (cGMP), is nitrated in the presence of nitric oxide, forming a 'second messenger' in the signal-transduction pathway: 8-nitrocGMP. The group also found that this second messenger chemically tags other proteins through a newly identified process known as *S*-guanylation.

MEDICINE

Every breath you take

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0706533104 (2007)

Monitoring blood sugar is a matter of life and death for diabetics, but invasive needle pricks are a burden, especially to children. In search of an alternative, Pietro Galassetti and Sherwood Rowland at the University of California, Irvine, and their colleagues analysed the breath of children with diabetes for exhaled indicators of blood sugar.

Using a very sensitive technique involving gas chromatography and mass spectrometry, the team measured 100 exhaled compounds in vanishingly small quantities as the patients received insulin to lower blood sugar. Changes in methyl nitrate levels were most closely correlated with alterations in blood sugar.

The team hopes to find a suite of such volatile organic compounds that could serve as a signature for other important blood markers, including insulin and lipids.

PHYSICS

Water that sticks

Phys. Rev. Lett. **99,** 148301 (2007)

What drives hydrophobic molecules to stick together when they are immersed in water? In 1945, scientists suggested that water forms an ice-like cage around hydrophobic groups. According to this theory, aggregation of the hydrophobic groups would be favoured because it would set the immobilized water free, increasing the system's entropy.

Yves Rezus and Huib Bakker of the FOMInstitute for Atomic and Molecular

Physics in Amsterdam, the Netherlands, find, with infrared spectroscopy, that water molecules around hydrophobic methyl groups on organic molecules rotate at least five times more slowly than they do in bulk water. Although not quite the ice cage originally proposed, this slowing down may help to explain the hydrophobic interaction.

NEUROSCIENCE

Hard to forget

Cell **131,** 160–173 (2007)

Researchers have unpicked the mechanism by which memories tied to strong emotions are recalled with greater clarity. The effect has been linked to release of the neurotransmitter noradrenaline during emotional situations. Now, Roberto Malinow of Cold Spring Harbor Laboratory in New York and his colleagues have determined that noradrenaline acts by regulating a class of receptor, known as GluR1-containing AMPA receptors, that is involved in learning.

In mice, both the fear caused by exposure to fox urine and the experimental injection of adrenaline, which boosts noradrenaline, triggered the addition of phosphates to two sites in GluR1. This phosphorylation meant that GluR1 could be more easily incorporated into synapses, which improved the animals' learning in behavioural tests carried out immediately after adrenaline injection. Noradrenaline had no effect on learning in mice that contained a mutant GluR1 that lacked the phosphorylation sites.

ANIMAL BEHAVIOUR

An ear to the ground

Biol. Lett. doi:10.1098/rsbl.2007.0443 (2007)

The marine iguanas of the Galapagos eavesdrop on the warning cries of

M. WIKELSKI

mockingbirds, report Maren Vitousek of Princeton University in New Jersey and her team. It's the first time that a non-vocal species has been spotted taking advantage of another animal's vocal communication.

Vitousek's team played recordings of different mockingbird calls to marine iguanas (*Amblyrhynchus cristatus*) on Santa Fe island, where both species are vulnerable to hawk attacks. In response to 'alarm' calls, about 45% of iguanas raised their heads or began to move away. Only 28% on average responded to playbacks of non-alarm birdsong. The tactic may give the iguanas — whose need to sunbathe can conflict with their need to keep watch — an early warning of the predators' presence. The picture (above) shows an iguana that didn't get away.

CHEMICAL BIOLOGY

Bittersweet reaction

Angew. Chem. Int. Edn **46,** 7697–7699 (2007)

As those who take artificial sweeteners will know, saccharin isn't all sweetness — it leaves a bitter metallic aftertaste. Although this seems to be harmless, no one knows what causes it. Now Gerhard Klebe of the Philipps University of Marburg in Germany and his co-workers think they have a clue.

They have found that saccharin is a surprisingly potent inhibitor of certain carbonic anhydrases — enzymes that have various roles in metabolism, acidity regulation and other aspects of biochemistry. Nanomolar concentrations of saccharin deactivate several variants of the enzyme, including one present in saliva that is involved in olfaction and taste. Compounds used to inhibit these enzymes for clinical purposes also have bitter, metallic aftertastes, so this might be a general side effect of such inhibitory action.

SYSTEMS BIOLOGY

Circadian *ménage à trois*

Science doi:10.1126/science.1148596 (2007)

Most organisms keep time with a circadian clock driven by changes in the expression of particular genes, but cyanobacteria do it differently. Erin O'Shea at Harvard University in Cambridge, Massachusetts, and her colleagues have devised a model that could explain how they tick.

The clock of the cyanobacterium *Synechococcus elongatus* centres on a protein, KaiC, that has different phosphorylation states. Over a roughly 24-hour period, a protein called KaiA functions with KaiC to add phosphates to two sites on KaiC, until another protein, KaiB, joins the complex and blocks KaiA's action. The phosphates then fall off.

Shea's team noticed an order to how the phosphates add and drop off that determines when KaiB can bind KaiC. This detail allowed them to explain how collections of these proteins maintain a cycle, even in a test-tube, instead of sitting in equilibrium.

JOURNAL CLUB

Francis Albarede
Ecole Normale Supérieure de Lyon, France

A geochemist goes à la recherche des climats perdus.

As a young postdoc at the California Institute of Technology (Caltech) in Pasadena I remember glancing through the 1952 logbook of a gas mass spectrometer while the machine readied my samples. In the book, Sam Epstein, one of the founders of modern geochemistry, had scribbled numbers representing the first attempt to determine past temperatures from oxygen-isotope abundances in fossils.

Since Epstein's measurements, the abundance of oxygen-18 in the carbonate skeletons of fossil sea creatures has become a broadly used indicator of past ocean temperatures. Such data are key to understanding modern climate change. But the usefulness of ^{18}O in 'palaeothermometry' is limited by problems including variations in oxygen-isotope levels in sea water and in the way different organisms take up the isotopes.

Recently, a group at Caltech proposed a measurement that may work better. As before, the carbonates are broken down into carbon dioxide for analysis. Instead of looking only for molecules containing ^{18}O, the Caltech team measures the abundance of molecules that contain both ^{18}O and the uncommon carbon isotope, carbon-13. The excess of this species over what would be expected through random combination of carbon and oxygen atoms indicates the temperature at which the carbonate formed.

Early tests of this 'clumped' thermometer on corals and fish ear bones were promising (P. Ghosh *et al. Geochim. Cosmochim. Acta* **70,** 1439–1456; 2006; and **71,** 2736–2744; 2007). Since then, the method has provided a new record of ocean temperature during the Palaeozoic era, which began 543 million years ago (R. E. Came *et al. Nature* **449,** 198–201; 2007).

I believe that clumped isotope thermometry is going to be a valuable new tool for palaeoenvironmental studies.

Discuss these papers at **http://blogs.nature.com/nature/journalclub**

NEWS

Biologists claim Nobel prize with a knock-out

The architects of a technique that has allowed biologists to identify the function of genes easily have been rewarded for their efforts with this year's Nobel Prize in Physiology or Medicine.

The technique allows researchers to generate 'knock-out' mice —mutant strains in which specific genes are disabled. These can be used to establish what role specific genes have in health, development and disease, and to create animal models of human diseases.

"Virtually no field of biomedicine has been untouched by one knock-out strain or another in a significant way," says Jeremy Berg, director of the National Institute of General Medical Sciences in Bethesda, Maryland.

Mario Capecchi from the University of Utah in Salt Lake City, Martin Evans of Cardiff University in Wales and Oliver Smithies of the University of North Carolina at Chapel Hill, share the €1.1-million (US$1.5-million) award. As is now often the case with this Nobel prize, the trio's work had previously been recognized with a Lasker award, in 2001.

Thousands of strains of knock-out mice have been generated since use of the technique was first reported in 1989. More than 500 of these are models for specific human disorders such as cardiovascular and neurodegenerative diseases, and cancer.

The technology's origins lie in a natural phenomenon called homologous recombination, which cells are thought to exploit to repair damaged DNA. Chromosomes, which package DNA, exist in pairs — one inherited from each parent — and during homologous recombination fragments of DNA can be exchanged between the two. Capecchi and Smithies found that artificial DNA of known sequence could engage in homologous recombination with mouse DNA, and exploited this to target specific mouse genes.

Evans provided the key element of heritability that eventually led to the development of knock-out mice — strains in which a gene remains knocked out in future generations. He had the idea of using mouse embryonic stem cells to introduce genetic material into embryos from a different strain of mouse. When he injected the stem cells into the embryos, their chromosomes, as expected, combined. The mosaic embryos created could be brought to term in a surrogate mother. When the pups were mated, their offspring contained genes derived from the stem cells. Evans then began modifying the stem cells before injecting them into the mouse eggs, using retroviruses to integrate new genes into their genomes. These new genes could then be transferred into the embryos and their offspring. Combining this technique with artificial homologous recombination led to the development of the first knock-out mouse.

Important refinements to the technology — particularly the development of 'conditional mutants' — have made knock-outs even more valuable for biologists. A system, known as Cre-lox, has been developed in mice by Klaus Rajewsky, now of Harvard Medical School, to allow the targeted gene to be switched off at a chosen time after birth. This is important both because up to 15% of genes are essential for embryonic development and a knock-out would not survive to birth, and because some genes may become relevant for a particular disease only later in life.

Smithies told *Nature* that the prize was "not unexpected, given the calibre of the work". Celebrating with ice-cream, he confessed to "being pleased". What thrills him most, he says, is opening journals to find so many papers that rely on knock-out mice. "It's quite clear that the homologous-recombination method has contributed deeply to our understanding of the genome," he says. "We may have the sequence, but knock-outs help us understand how it functions."

In recent years, Capecchi has helped elaborate the role of particular genes in embryonic development. His work has focused particularly on body organs and the way the body plan is formed — vital to ensuring that all of an animal's body parts are in the correct places.

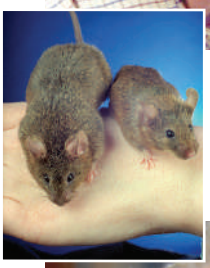
Evans has developed many mouse models of important human diseases, such as cystic fibrosis, and has used these to study disease mechanisms and to try to find ways to repair defective genes. Smithies has also developed a cystic fibrosis mouse as well as models of common diseases including high blood pressure and atherosclerosis.

Now that the mouse genome has been sequenced, a worldwide effort has been launched to knock out every single one of the animal's genes.

"The impact of the technology on the understanding of gene function and its benefits to mankind will continue to increase for many years to come," says Evans. ■

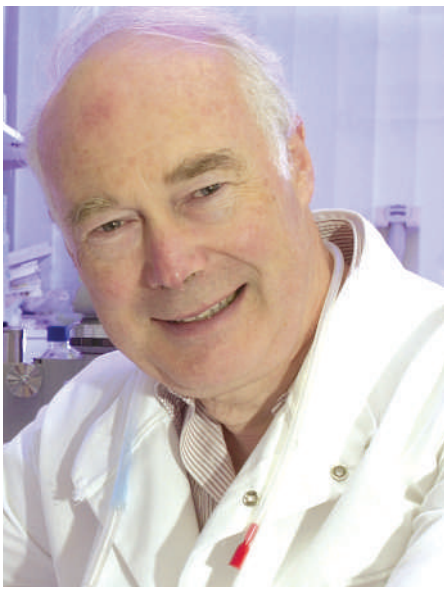
Alison Abbott

CARDIFF UNIV./AP/D. C. PIZAC/AP/K. TAM



AP/K. WELLER

Oliver Smithies, top, Mario Capecchi, middle, Martin Evans, bottom, and knock-out mice, left.



ALAMY

CHEMISTRY NOBEL
See online for coverage of the Nobel Prize in Chemistry.
www.nature.com/news

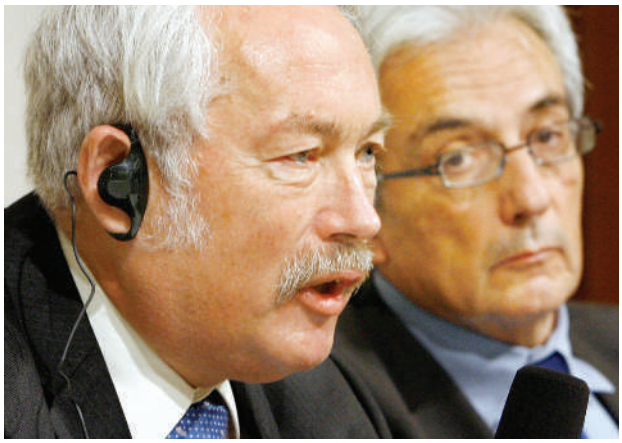
The physics prize inside the iPod

Two researchers who discovered an effect that has dramatically shrunk the size of magnetic storage devices have won the 2007 Nobel Prize in Physics.

Albert Fert of the University of Paris-South in France and Peter Grünberg of Jülich Research Centre in Germany split the prize for their 1988 discovery of an effect called giant magnetoresistance (GMR). The Royal Swedish Academy of Sciences announced the award on 9 October in Stockholm.

The effect has been heralded as one of the first major applications of the fields of nanotechnology and 'spintronics'.

"I am so proud and so happy," Fert said in a press conference via telephone from France. "Science is something marvellous."



T. YAMANAKA/AFP/GETTY IMAGES

Spin doctors: Peter Grünberg (left) and Albert Fert.

At the heart of GMR are the spins of electrons, which generate a magnetic field and can be aligned either up or down. An electron can easily pass through a material whose electrons are similarly aligned, but will encounter resistance when it passes through one with electrons aligned in the opposite direction.

Fert and Grünberg discovered the effect independently of each other using multiple layers of magnetic and non-magnetic materials only tens of nanometres thick. When all the layers were aligned in the same direction, say 'up', electrons with the same alignment passed through the material easily, whereas those with the opposite alignment struggled. But when the layers were organized in an alternating 'up-down' alignment, all electrons encountered resistance. The net effect was a rise in resistance that was much bigger than any seen before — hence 'giant'.

This led to devices that are very sensitive to tiny magnetic fields. A hard disc drive stores bits on its surface as a pattern of magnetic fields. Until the discovery of GMR, hard discs used metal induction coils to read out the data. But the laws of induction meant that the coils, and thus the bits, had to be quite large. GMR opened up a way to build much smaller magnetic heads, says Claude Chappert of the University of Paris-South. The discovery revolutionized consumer electronics. "I think this triggered the common use of MP3 players," he notes.

By the end of the 1990s, the technology had become standard across the electronics industry, thanks partly to the work of physicist Stuart Parkin at IBM's Almaden Research Center in San Jose, California, who came up with a simple way to create the thin multilayers. Although Parkin has shared physics prizes for GMR with Fert and Grünberg in the past, he was not included in the Nobel announcement. Parkin conducted vital work that allowed the effect to be commercialized, but Fert and Grünberg were the ones who discovered it, says Tony Bland, from the University of Cambridge, UK. "I think the field will generally see this as fair," he adds.

Storing information is not the only application, says Fert. The discovery has also opened the door to the possibility of 'spintronics', the idea of using electrons' spins, as well as their charge, in electronic devices. Spintronics could soon lead to random-access memory that remains stable even without power, securing data and allowing some computers to start up more quickly, says Bart van Wees of Groningen University in the Netherlands. It could also create new ways for fibre-optic systems and conventional semiconductors to talk to each other.

Even further out is the possibility of processing information using spin, rather than electrical current. Although still highly speculative, Chappert says, such a computer could run faster and on much lower power than existing devices. "Spintronics could bring a lot." ■

Geoff Brumfiel

Immigration institute provokes outcry from social scientists

The French government is to create a powerful institute for research on immigration and integration. The move has sparked opposition among social scientists, who claim that the body is a thinly veiled bid to exercise political control over their research.

The Paris-based institute will be placed under the High Council for Integration, which is affiliated with the prime minister's office. It is expected to be inaugurated shortly by Brice Hortefeux, the immigration minister.

The government says that the 23-member institute will serve as a 'one-stop shop' for distributing both public and private research funds, and help define "pertinent fields and topics" for study.

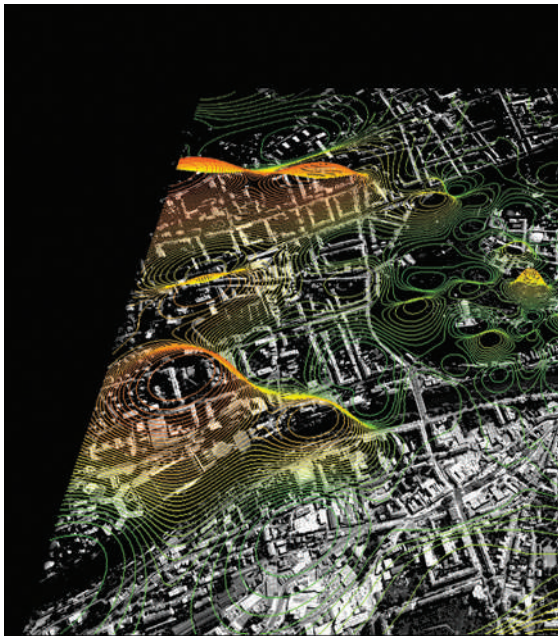
But in a statement issued on 2 October, 75 French historians and intellectuals expressed their "grave concerns" over the institute, branding it a "threat to academic freedom". That threat arises, they say, "in a context where the political discourse tends more-and-more to present immigration as a danger, where successive legislation has increasingly restricted the rights of foreigners and where rhetoric serves to mask discrimination". A tough immigration bill is currently going through the French legislature (see *Nature* 449, 377–378; 2007).

The dissenters also criticized the nomination of Hélène Carrère d'Encausse, a specialist in Russian affairs and secretary of the French Academy, as the institute's president. During urban riots in 2005, Carrère d'Encausse sparked controversy by identifying polygamy among immigrants as one of the causes.

Ironically, several of the statement's signatories are listed as members of the new institute in a document released by the government on 19 September. Paul Schor, a specialist in American minorities at the School of High Studies in Social Sciences in Paris, says that he first learnt of his inclusion when the list was circulated on the Internet. He says he has had no explanation for being included in the list, and that it was done without consultation. ■

Declan Butler

Data sharing threatens privacy



When web provider AOL's research division published an analysis of search behaviour on the Internet last year[1], it had what it thought was a bright idea: it would reach out to academics by making an anonymized version of the data freely available for download from its website. But within hours, it had to pull the site, after bloggers managed to infer many identities from the data and view the associated search histories.

AOL's mistake highlights a growing dilemma for the social sciences. The hottest growth area in the field is computational social science. This is often based on privileged access to electronic data sets such as e-mail records, mobile-phone call logs and web-search histories of millions of individuals.

Such studies are ushering in a revolution in the social sciences, specialists say. But there is a trade-off between the scientific interest in working with such data and concerns about privacy. "It's a huge issue," says David Lazer, a researcher at the John F. Kennedy School of Government at Harvard University.

Lazer is co-author of a social-network analysis based on the largest mobile-phone data set ever studied by academics — 18 weeks of details of who called whom, when and for how long, among 7 million users, representing 20% of the population of an unidentified European country, supplied by an unnamed operator[2]. A month after it appeared, Microsoft researchers published an even larger study[3], this time of instant messages, featuring 30 billion conversations among 240 million people worldwide.

Until now, social science has struggled to obtain tools that do more than scratch the surface of some of its questions. These range from identifying the driving forces behind violence, to the factors influencing how ideas, attitudes and prejudices spread through human populations. The available tools have largely remained in a time warp, consisting of analyses of national censuses, small-scale surveys, or lone researchers with a notebook observing interactions within small groups.

Being able to automatically and remotely obtain such massive amounts of continuous data opens up unprecedented opportunities for social scientists to study organizations and entire communities or populations, says Marshall Van Alstyne, a researcher at Boston University. He is conducting research on the dynamics and productivity of organizations by analysing network patterns of e-mails among volunteers.

"There is enormous potential here for lines of research that shed new light on basic social-science questions," says Jon Kleinberg, a specialist in network analysis at Cornell University in Ithaca, New York.

But the privacy issue looms large. Repetition of the AOL gaffe by other researchers might create a damaging public backlash, warns Myron Gutmann, director of the Inter-university Consortium for Political and Social Research, based in Ann Arbor, Michigan. The consortium acts as a clearinghouse for secure access to conventional sensitive social-science data sets, such as the raw data of the US national census.

"So far, researchers have been careful, and successful at avoiding serious problems," says Kleinberg. "But as the number of these types of study increases, the community is clearly going to need to engage in deeper discussions about the right way to safeguard privacy in working with these kinds of data."

Lazer says the mobile-phone study was possible only after taking many precautions, including anonymizing data before researchers gained access to them, confidentiality agreements between the phone company and Harvard University, and approval and stipulations on access controls by the university's internal review board.

But those ad hoc arrangements can go only so far, says Lazer, and a larger institutional framework is needed to set best practices, especially for safe sharing of data among academics. The social scientists pioneering such research often have computer-science backgrounds. As the software tools developed go mainstream, they get taken up by academics with less expertise in protecting data from abuse.

Although AOL had anonymized the data it released — 20 million web queries from around

Doctors not to blame over HIV infection by tainted blood

A. WYLD/CANADIAN PRESS/AP

Acquitted: Roger Perrault leaves court.

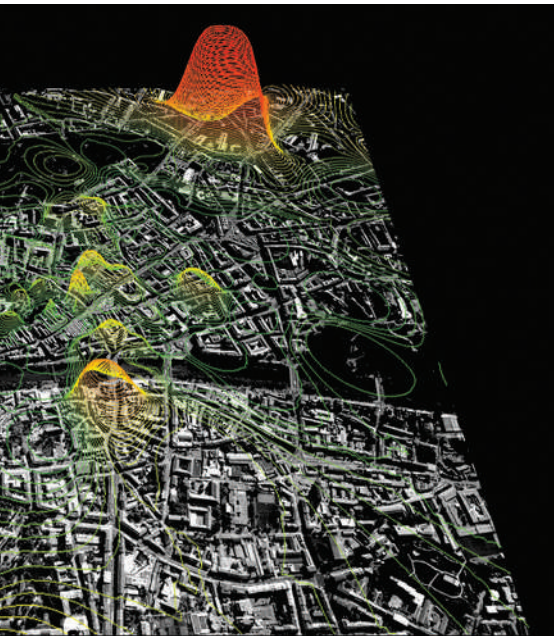
A Canadian court has acquitted four doctors and a US blood products company of criminal negligence in the case of four haemophiliacs who were infected with HIV after receiving transfusions of tainted blood in the 1980s.

After a five-year police investigation and a lengthy trial that involved more than 100 witnesses and 1,000 exhibits, Judge Mary Lou Benotto of the Ontario Superior Court of Justice in Toronto effectively said that the doctors and the company were not only acquitted but fully exonerated.

"The allegations of criminal conduct on the part of these men and this corporation were not only unsupported by the evidence, they were disproved," Benotto wrote in her 1 October decision. "The events here were tragic. However, to assign blame where none exists is to compound the tragedy."

The case is the latest of numerous global investigations into the circumstances in which thousands of patients were given infected blood even after it became known, in autumn 1984, that heat treatment killed HIV in blood products. More than 1,000 Canadians, about 700 of them with haemophilia, were infected by HIV from transfusions, almost all of them before mid-1985.

The Canadian court case



MIT SENSEABLE CITY LAB.

The data behind analyses, such as this map of mobile-phone use in Rome, present privacy issues.

650,000 users over 3 months — the company overlooked the fact that people's search queries often include vanity searches on their own name. This fact, combined with context from other search queries, often allowed complete search records to be narrowed down to individuals.

Using algorithms similar to those for cracking cryptography keys, Kleinberg also showed in an May paper[4] that more sophisticated attacks can be used to identify people in completely anonymized data. In the model, individuals in the network, and simulated false users, were able first to identify themselves in the network, and subsequently obtain the supposedly confidential data of all those connected to them. "It's hard to guarantee in practice that any data set, however well anonymized, is truly safe from privacy breaches," says Kleinberg.

This work reinforces the need for a systematic, institutional approach to improving the privacy rights of those whose data are used, says Van Alstyne. That echoes the conclusions of a May study by the US National Academies[5], which said that safeguarding privacy cannot safely be left to individual researchers. It stated that: "Institutional solutions involve establishing tiers of risk and access, and developing data-sharing protocols that match the level of access to the risks and benefits of the planned research."

But Gutmann and other social scientists also stress that the risks should be kept in perspective. Scientists must meet strict rules on any research on human subjects. In contrast, private firms are largely free from such constraints, and already have wide latitude to snoop on, and data mine, their employees' work habits.

Specialized firms, not to mention governments, are also building vast cross-referenced databases of information on every aspect of individuals' lives, including their web search behaviour, interests and personal preferences. An excessive focus on research would be "fixing the leaky faucet when the bath-tub's overflowing", says Van Alstyne. "Businesses seem more prone to misuse private data than scientists of any stripe." ■

Declan Butler

1. Pass, G., Chowdhury, A. & Torgeson, C. *Proc. 1st Int. Conf. Scalable Information Sys.* doi:10.1145/1146847.1146848 (ACM Press, New York, 2006).
2. Onnela, J.-P. *et al. Proc. Natl Acad. Sci. USA* **104,** 7332–7336 (2007).
3. Leskove, J. & Horvitz, E. *Microsoft Res. Tech. Rep.* MSR-TR-2006-186 (2007).
4. Bachstrom, L., Dwork, C. & Kleinberg, J. *Proc. 16th Int. World Wide Web Conf.* doi:10.1145/1242572.124259 (2007).
5. *Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data* www.nap.edu/catalog/11865.html (Natl Acad. Sci., Washington DC, 2007).

See Editorial, page 637.

addressed the infection of 4 people in British Columbia and Alberta in 1986 and 1987. They received an HIV-infected blood-clotting product made by Armour Pharmaceutical, then a maker of blood products based in New Jersey.

The now-elderly doctors charged were Roger Perrault, then the national director of the Canadian Red Cross blood transfusion service; John Furesz, then director of Health Canada's Bureau of Biologics (BDB); Wark Boucher, who headed the BDB's blood product division; and Michael Rodell, then a vice-president of Armour Pharmaceutical.

The Canadian Hemophilia Society said it was "surprised and disappointed" by the outcome. "This verdict sends the wrong message to those responsible for the health of the public," Pam Wilton, its president, said in statement.

But David Scott, an Ottawa lawyer who defended Furesz, said: "These charges were undoubtedly politically motivated. If those in charge of the criminal process succumb to political pressure to lay criminal charges in inappropriate cases, they will unwittingly undermine the public confidence in the administration of justice."

Lawyers for the Crown didn't say whether they plan to appeal within the 30-day time frame allowed. "I cannot imagine them doing so," says Edward Greenspan, the Toronto lawyer who represented Perrault. "But then again I couldn't imagine them spending 17 months in a court of law proving nothing." ■

Meredith Wadman

ZOO NEWS

Redhead goes for blondes
Sibu the orang-utan is driving staff at Apenheul Primate Park in the Netherlands crazy by refusing potential mates. Staff report that the 31-year-old male seems to be interested only in tattooed blonde women, having taken quite a shine to one of his previous keepers.

SCORECARD

Peaceful sleep
Partners of heavy snorers can breathe a sigh of relief — a new computerized pillow senses the sound of snoring and inflates or deflates to alter the angle of the slumberer's head to allow easier breathing.

Concussion
Also new to the technology market is a computerized helmet for American football, which analyses impacts and creates a profile for each player, tallying up the inevitable carnage and warning of impending injury.

SHOWBIZ NEWS

Star (Trek) in the sky
George Takei (pictured), perhaps better known as USS Enterprise helmsman Mr Sulu, is the latest sci-fi icon to have an asteroid named in his honour by the International Astronomical Union. The rock, now called 7307 Takei, joins asteroids named after *Star Trek* creator Gene Roddenberry and Nichelle Nichols, aka Lieutenant Uhura.

Sources: Associated Press, Reuters, Scientific American

E. FIORE/REX FEATURES

SIDELINES

Congress grills officials over biosafety boom

Critics are questioning whether a multi-billion-dollar programme to improve public safeguards against dangerous pathogens is really making the United States safer than it was before.

The criticism surfaced in Washington on 4 October, when members of Congress grilled officials of the Bush administration over the rapid expansion of laboratories that house pathogens such as anthrax and Ebola. The same day, the Government Accountability Office, an investigative arm of Congress, reported that the government wasn't keeping tabs on the number of such labs in operation, or coordinating their regulation.

"Are so many labs doing this research that you actually increase the chances of a catastrophic release of a deadly disease?" asked Bart Stupak (Democrat, Michigan) at the hearing of the House energy and commerce committee. The committee heard testimony from senior officials of the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC), the agencies that oversee much of the work.

The hearing came three months after the CDC halted research on potential bioweapons agents at Texas A&M University in College Station, when a lab worker there contracted brucellosis (see *Nature* **448,** 105–106; 2007).

Documents uncovered by Associated Press have since identified more than 100 mishaps ranging from bird flu transmitted to a lab worker by a ferret bite, to missing vials of anthrax.

Critics of the expansion claim that the mishaps indict the entire biosafety research programme. "This work cannot be safely performed in a university or corporate setting," says Richard Ebright, a molecular biologist at Rutgers University in Piscataway, New Jersey.

Before the terrorist attacks of 11 September 2001 and the subsequent anthrax releases that killed five people in the United States, only a few studies of dangerous pathogens were done outside fortress-like government facilities such as the CDC's laboratories in Atlanta, Georgia, and the Army Medical Research Institute of Infectious Diseases in Fort Detrick, Maryland.

But to broaden its knowledge and engage leading scientists, the Bush administration ramped up biosecurity research outside those facilities. In 2001, the National Institute of Allergy and Infectious Diseases — the NIH institute that has led the biosecurity agenda — spent $41 million on such work. Last year, it spent $1.6 billion. And the number of biosafety 4 laboratories studying deadly pathogens without cures has grown from 5 in 2001 to 15 either in operation or being planned today. Currently, 409 laboratories staffed by 14,000 people tinker with the various 'select agents' tracked by the CDC.

So far, none of the security lapses is known to have harmed the public, says microbiologist Ronald Atlas, graduate dean at the University of Louisville in Kentucky. But Atlas worries that the government has placed more emphasis on physical security than on training people to handle pathogens safely. "When you work in a lab, you can become complacent," he admits, "and that just can't be allowed today." But he is against the scaling back of pathogen research.

Europe ponders restrictions on life sciences

European biologists have criticized the idea of a 'two-tier' research publication system floated by the European Commission.

In a 'green paper' published on 11 July, the commission asked whether sensitive biological research should be published twice: a public version without sensitive content and a full-content one for "relevant bio-stakeholders". But several groups have taken issue with the ideas proposed in the discussion document.

The European Biosafety Association (EBSA), which represents people working in biosafety and associated activities, said that it thought the two-tier approach was "unworkable and contrary to scientific freedom".

Similarly, the bioindustry body EuropaBio, based in Brussels, Belgium, said: "Pre-publication review of scientific research and possibly classification of its release raises profound questions for scientific freedom — questions that have been addressed in the nuclear field but not substantially in the life sciences."

"In principle it sounds good, but who is going to be the gatekeeper?"

Neither group spelled out how it thinks that information on potentially dangerous pathogens, for instance, should be contained. The EBSA, however, suggests that professional associations should develop guidelines on this matter.

The paper provides no details on how a two-tier publication system might work in practice. John Marks, chief executive of the European Science Foundation, a coalition of European research agencies based in Strasbourg, France, says that the question of who might operate such a system remains unresolved. "In principle it sounds good," he says. "But the question is: who is going to be the gatekeeper?"

The still-unexplained anthrax attacks in the United States in 2001, together with publications such as those describing the creation of a particularly virulent version of mousepox[1] and the synthesis of an influenza strain containing parts of the 1918 pandemic virus[2], has led to the consideration of various approaches to control access to some types of biological information (see *Nature* 441, 388–389; 2006). The green paper was a part of this process, and together with the comments received on it, could eventually lead to European legislation.

As well as the two-tier system of publication, the green paper suggests that

T. HOROWITZ/CORBIS

Could investment in biosafety research be leaving the public more at risk?

"We need more infectious-disease research, not less," he says.

Despite reservations expressed at the hearing, the expansion of bioterror-related research continues. Congress might still use its control of budgets to push for greater oversight, though, and Stupak and others have asked the government to assess the overall programme. ■

Ewen Callaway

research grants should be awarded, in part, on the basis of applicants' compliance with biosafety standards. It also puts forward a Europe-wide system for the accreditation of research facilities suitable for work on potentially dangerous pathogens. And it asks whether it would be feasible to accredit individual researchers.

In their responses, the EBSA and EuropaBio backed the accreditation of facilities but said that it would be much more difficult to accredit individual researchers without running into problems regarding discrimination or personal privacy. Both groups broadly welcomed the idea of making grants conditional on compliance with biosafety standards. ■

Daniel Cressey

1. Jackson, R. J. *et al.* A. *J. Virol.* **75,** 1205–1210 (2001).
2. Tumpey, T. M. *et al*. *Science* **310,** 77–80 (2005).

Doubts raised over stem-cell marker

Two papers published this week challenge the validity of using the protein Oct4 as a marker for adult stem cells. The protein maintains embryonic stem cells in an undifferentiated state, called pluripotency, but the papers question whether it has the same function in adult stem cells.

Oct4 controls gene expression during the early stages of embryo formation, but switches off in somatic cells shortly thereafter. It thus seems logical that Oct4 might also control adult stem cells, and more than 60 studies have used Oct4 expression as a marker of these cells.

But now, Rudolf Jaenisch's lab at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, has used several approaches to test for Oct4 expression in adult stem cells in mice[1]. All came up negative. "It's hard to prove beyond any doubt that something is not expressed or doesn't have a function," says author Christopher Lengner, also from the Whitehead Institute. "That's why we had to use overlapping methods. We beat the thing to death."

Lengner and his colleagues aren't alone. Several years ago, Lorenza Lazzari, of the Maggiore Foundation Hospital in Milan, Italy, and her colleagues detected low levels of Oct4 expression in adult stem cells, but their negative controls, which did not contain stem cells, yielded the same result.

Lazzari had found it difficult to publish her work, but eventually succeeded earlier this year[2]. "The dogma of Oct4 as a stem-cell marker was stated by several scientists working on somatic stem cells and it has been very difficult to convince editors," says Lazzari, who says that she has had several manuscripts on the topic rejected. Many published papers showing Oct4 expression simply didn't include a negative control, Lengner says.

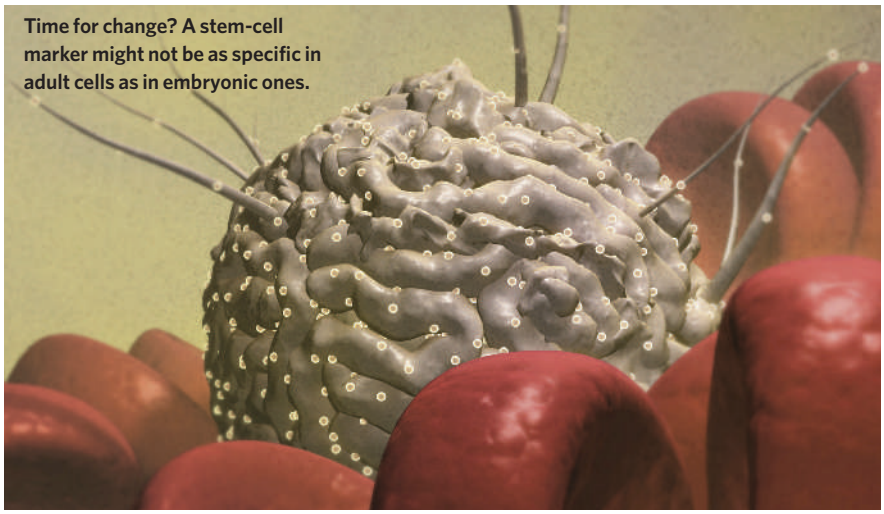
Lazzari's false-positive results could be due to the presence of several *Oct4*-related genes that do not code for functional protein, says Gesine Kögler of the Heinrich-Heine University in Düsseldorf, Germany. Kögler and her colleagues found that many approaches used to detect Oct4 cannot distinguish between the functional *Oct4* gene and its related pseudogenes[3]. Lengner also says that many papers relied on stem cells that had been maintained as cell cultures for many generations, and may no longer reflect what occurs in live animals.

This doesn't necessarily mean that adult stem cells are not pluripotent, notes Leendert Looijenga, a pathologist at the Erasmus Medical Center in Rotterdam, the Netherlands. It does, however, suggest that they are regulated differently from embryonic ones.

Looijenga and his colleagues have tested more than 3,600 cancer tissue samples, representing more than 100 cancers. They found that Oct4 was often expressed in gamete-producing tissue, but not in other types of tissue, supporting the link between pluripotency and the protein. Researchers were driven by the idea that Oct4 ought to be present in adult stem cells and stem-cell-derived cancers, says Looijenga. "It's a logical idea," he says, "but logical ideas aren't always right." ■

Heidi Ledford

1. Lengner, C. J. *et al. Cell Stem Cell* **1,** 403–415 (2007).
2. Zangrossi, S. *et al. Stem Cells* **25,** 1675–1680 (2007).
3. Liedtke, S., Enczmann, J., Waclawczyk, S., Wernet, P. & Kögler, G. *Cell Stem Cell* **1,** 364–366 (2007).
4. Looijenga, L. H. J. *et al. Cancer Res* **63,** 2244–2250 (2003).



Time for change? A stem-cell marker might not be as specific in adult cells as in embryonic ones.

MEDICALRF.COM/CORBIS

Weapons of war scoop spoof Nobels

It was a case of 'make love not war' winning out over 'don't ask, don't tell'. Borrowing a bad story line from a B movie, researchers at the US Air Force Research Laboratory at the Wright-Patterson Air Force base in Ohio proposed to develop chemical aphrodisiacs that would make enemy soldiers sexually irresistible to each other. The idea — dubbed the 'gay bomb' — earned the unnamed Ohio scientists the 2007 Ig Nobel Peace Prize, which was awarded at Harvard University on 4 October along with nine other prizes.

The competition for the Peace Prize was intense even within Wright-Patterson laboratory itself. Researchers there also proposed chemicals designed to attract stinging and biting creatures, as well as low-toxicity compounds that would "create severe and lasting halitosis" for those exposed to tiny concentrations. Despite the merits of the competing concepts, none had quite the 'legs', so to speak, of the gay bomb.

The Ig Nobels — a spoof of the real Nobel prizes (see page 642) — have been awarded annually since 1991. But the gay bomb would never have basked in their glow had it not been for the Sunshine Project in Austin, Texas. The group's director Edward Hammond caught wind of the report when he saw a promotional CD-ROM of 'non-lethal weapons', prompting him to seek all supporting documents under the Freedom of Information Act.

This year's Ig Nobel prize for medicine went to Brian Witcombe (left) and Dan Meyer for their finding that sword-swallowing can damage your health.

The fate of the gay bomb project is unclear, says Ig Nobel impresario Marc Abrahams, who was told by former lab staff that "if research on the gay bomb had proceeded further, it would have immediately been stamped secret, and neither you nor I would have any way of knowing about it." Fortunately, however, the Ig Nobel selection committee was lucky enough to hear about the idea, as well as others that warranted similar distinction.

The Linguistics Prize, for example, went to researchers from the University of Barcelona in Spain for showing that rats cannot always distinguish between a person speaking Japanese backwards and a person speaking Dutch backwards.

Brian Witcombe of Gloucestershire Royal NHS Foundation Trust, UK, shared the Medicine Prize with Dan Meyer of Sword Swallowers' Association International in Antioch, Tennessee, for their report, "Sword Swallowing and Its Side Effects". In recommending the paper for the *British Medical Journal*, one reviewer called it "a cut above the others", according to Witcombe.

Investigators at the National University of Quilmes in Argentina won the Aviation Prize for determining that Viagra aids hamsters' recovery from jet lag. A member of that team, Diego Golombek, praised his graduate students for top-flight research and "for going to the drugstore to get the Viagra". ■

Steve Nadis

Panel issues warning over mineral market tremors

The United States should establish an agency to track markets for minerals that are critical components of industrial products, says a panel of the National Academies.

In a report published on 5 October, the panel says that the existing Minerals Resources Program at the US Geological Survey in Reston, Virginia, does not have the resources, authority or autonomy needed to track the supply and demand of valuable minerals used in everything from televisions to cars. It cites the Energy Information Administration (EIA) — set up in Washington DC in the wake of the 1973 oil shock — as a model for a future federal agency.

"We need quality information to be able to assess the vulnerability to disruption," says Roderick Eggert, chair of the panel and a mineral economist at the Colorado School of Mines in Golden.

The report points out that the sharp fluctuations in mineral prices and the minerals' varied uses receive little public attention. Engineers have folded obscure minerals from all over the periodic table into high-tech alloys, films and catalysts that industry relies on, it notes.

J. L. AMOS/CORBIS

Platinum supplies are at risk.

And mineral markets have been rocked by wild price swings: in late 2002, for example, the price of indium — used in flat-panel displays — hit a historic low of $60 per kilogram. In 2005, the price reached $1,000 per kilogram (see *Nature* 449, 131; 2007).

Booming demand for minerals in India and China suggests that historical trends might be an unreliable guide for the next few decades, says Thomas Graedel, another of the report's authors and an industrial ecologist at Yale University in Connecticut.

The report identifies indium, manganese, niobium, platinum-group metals and rare-earth metals as the important minerals whose supply is most in doubt. The Minerals Resources Program has only 135 staff, compared with 350 at the EIA. Unlike the EIA, it has no legislative authority to demand market information from industrial companies and no right to publish data independently of its parent agency.

But Jerry Taylor of the Cato Institute, a pro-free-market think-tank based in Washington DC, says that replicating the EIA — which he brands a "subsidy to the oil industry" — is unnecessary. Prices already provide incentives for companies to pay for and gather mineral data, he says, and the risks of supply shocks are already factored into prices. ■

Eric Hand

University of California fined over security breach

The US Department of Energy has fined the University of California $3 million for violations related to a security breach at Los Alamos National Laboratory in New Mexico.

In a raid in 2006, police found classified material in the home of a subcontractor who had been scanning lab documents for a work project. The department determined that the University of California's security measures were "deficient" and partly to blame for the violation.

Los Alamos has been hit by a series of security lapses in recent years, causing the energy department to sever an exclusive contract with the University of California that dated back to the Second World War. Since June 2006 the university has managed the lab as part of a broader business consortium.

Spokesman Chris Harrington says that the university is reviewing the final notice and has not yet determined whether to appeal against the ruling.

Blueprint drawn up for Britain's science spending

Britain needs to invest more in high-tech research and innovation if it is to remain competitive in a globalized economy, says its former science minister. David Sainsbury makes his recommendations in a government-commissioned review of British science and innovation.

The report, entitled *The Race to the Top*, calls for Britain's Technology Strategy Board to oversee government-funded research councils and make it easier for industry to gain access to research and development funds, with the aim of fostering technological innovation. "Company strategies based on low costs alone will end up in a downward spiral," Sainsbury says.

The recommendations in the report, released on 5 October, will provide a strategic outline for a comprehensive overhaul of government science and other spending announced on 9 October. As *Nature* went to press, this spending review was expected to include a substantial boost in government funding for research.

Council of Europe votes against creationist teaching

Creationism is a potential threat to human rights and any attempts to incorporate it into science must be resisted, says the Council of Europe. The council is an intergovernmental body that is responsible for, among other things, the European Convention on Human Rights.

On 4 October, the council's Parliamentary Assembly voted in favour of its member states promoting evolution as "a fundamental scientific theory" and to "firmly oppose the teaching of creationism as a scientific discipline on an equal footing with the theory of evolution". The motion is not binding on the council's 47 member states.

Concerns have been increasing about the promotion of creationism in Europe. The council highlighted a number of campaigns in countries such as Britain, Turkey, France and Russia (see *Nature* 444, 406–407; 2006).

Conservationists spot flaws in plan to save owl

Concerned scientists and Democratic congressmen each sent their own letter to the US Department of the Interior on 2 October, protesting against the agency's draft plan to save the northern spotted owl (*Strix occidentalis caurina*) from extinction.

Conservation icon: the northern spotted owl.

J. SARTORE/NATIONAL GEOGRAPHIC/GETTY

The rescue plan for the conservation icon rests heavily on controlling another owl — the barred owl (*Strix varia*) — that outcompetes its spotted cousin. But the Society for Conservation Biology and the American Ornithologists' Union, both hired as peer reviewers by the government, as well as three owl experts who were also consulted, all felt that this emphasis was strange. They say that habitat loss due to logging is the clear cause of the owl's decline.

Meanwhile, the Government Accountability Office has begun examining several decisions based on the Endangered Species Act that have been criticized by environmentalists, including that on the spotted owl.

India gets high-security lab for human diseases

India will set up its first high-security facility for handling and doing research with highly infectious organisms that cause diseases in humans.

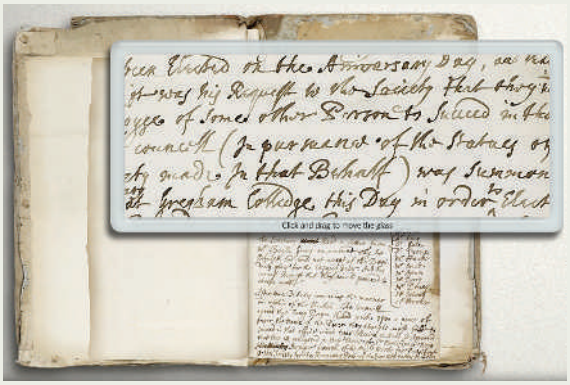
The US$25-million biosafety level-4 (BSL-4) lab is to be established on a 2-hectare site in Hyderabad, about 3 kilometres from the Centre for Cellular and Molecular Biology (CCMB). The centre has been given responsibility for setting up the new lab as a national facility.

India already has a BSL-4 lab for animal pathogens in Bhopal, and only last month validated and cleared BSL-3-level labs at five institutions. Details of containment facilities at the defence department's laboratory in Gwalior are not available.

Lalji Singh, director of the CCMB, says the new lab will not do any classified defence work: "The main objective of our facility would be to carry out basic research on the biology of lethal and highly infectious microorganisms."

Correction
The Business story 'The carbon cycle ride' (*Nature* **449,** 522–523; 2007) should have said that India has more UN-registered carbon-trading projects than China and Brazil combined — not that it has more than China and India combined.

Hooke online offers minute details of Royal Society



ROYAL SOCIETY

The original pages are too delicate to touch, but now anyone with Internet access can read the writings of Robert Hooke, one of seventeenth-century Britain's greatest natural philosophers. The Royal Society, of which Hooke was appointed secretary in 1677, has created an online version of the £1-million (US$2-million) notebook that it dramatically rescued from auction in March last year (see *Nature* **440,** 725; 2006).

The manuscript contains Hooke's personal minutes of Royal Society meetings from 1661 to 1691, and provides a glimpse of the scientific discoveries of the age, as well as Hooke's personal spats with other greats such as Isaac Newton. The online version of the manuscript (pictured) can be seen at www.royalsoc.ac.uk.

BUSINESS

Patent crunch pending

Drug makers are fighting a rear-guard action against patent laws that are before the US Congress. **Meredith Wadman** reports.

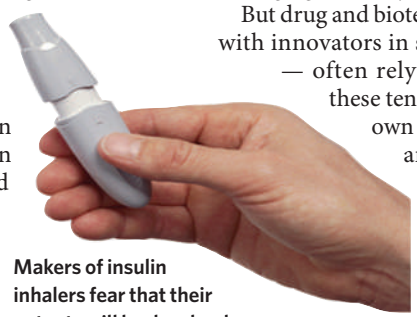
The biotechnology firm Alkermes in Cambridge, Massachusetts, has made a big bet on a simple idea: that patients with diabetes should be able to inhale, rather than inject, their insulin. Since the method was first described in 1997 (D. A. Edwards *et al. Science* **276,** 1868–1872), the company has been developing it to allow large drug particles such as insulin to be delivered efficiently from a palm-sized inhaler. With its product now in the final stages of clinical trials, Alkermes — in partnership with drug maker Eli Lilly — has poured hundreds of millions of dollars into the venture.

C. CORDER/NEWSCOM

"You shouldn't have to throw the biotech industry under a bus to make life easier for the IT industry." — Jim Greenwood

But Alkermes has a problem on Capitol Hill: a package of reforms that the company, along with the rest of the biotechnology and pharmaceutical industries, says threatens to undermine its patents. The legislation passed the House of Representatives last month on a closer-than-expected vote; it may come to a Senate vote as soon as this month.

The Patent Reform Act of 2007 is predicated on the view that the US patent system has swung too far in favour of innovators, who are collecting outsized damage awards, to the detriment of sometimes innocent infringers. A commonly cited example is the $1.52 billion levied against Microsoft in February by a jury in San Diego. The jury decided that the software giant had infringed two patents owned by Alcatel–Lucent, based in Paris, when it used the MP3 format in its Windows Media Player software. (The verdict was overturned in August by a judge who contended that only one patent had been violated. A new trial to calculate the damages is awaited.)



Makers of insulin inhalers fear that their patents will be devalued.

Damage limitation

Most controversial, is the part of the reform that aims to reduce the amount of damages paid to firms when patents are infringed. It tells the courts to base damage awards on the economic value of the patented component that was infringed, not on the value of the entire product, as often happens under current law. It also instructs the courts to subtract out the economic value of all 'prior art' — that is, any publicly available information relevant to claims of originality — in the invention. That's music to the ears of the Apples and Microsofts of this world, whose products often rely on thousands of components, many of which have been patented by others — and sometimes without the infringing company's knowledge.

But drug and biotechnology firms — along with innovators in several other industries — often rely on fewer patents, and these tend to be the result of their own original work. Thus, they are more likely to sue for infringement than be sued. What's more, what look like minor advances on prior art in the drug industry can yield big advantages to patients. Consider, for instance, that neither insulin nor inhalers are new. So subtracting out prior art in the case of inhaled insulin could mean that Alkermes is left holding little or nothing in damages if it were to sue another firm for infringement, the company says.

Kathryn Biberstein, Alkermes' general counsel, warned the Senate judiciary committee back on 6 June that the legislation would discourage companies from developing products such as inhaled insulin by removing their incentives to do so.

The upshot is that the proposed change in damage calculations is strongly opposed by drug and biotech companies — along with big manufacturing firms such as Caterpillar and General Electric, who preside over patent estates that include everything from batteries to jet engines. They argue that the new approach will devalue their patents, their licences and ultimately their attractiveness to investors.

"If your protein is only 5% different from the preceding protein, which was prior art, you will receive pennies on the dollar," says Jim Greenwood, who heads the Biotechnology Industry Organization in Washington DC. "Under that kind of approach, it pays to infringe. And if it pays to infringe, people will no longer invest in companies with novel therapeutic patents."

"What Congress needs to do is improve the patent environment for all sectors of the American economy," says Greenwood. "You shouldn't have to throw the biotech industry under a bus to make life easier for the IT industry."

But the bill's supporters, including most of the finance and IT industries, counter that the changes are vital to curb damages that have run amok.

"I had three pieces of patent litigation in the late 1990s. Today I am defending 30," Mark Chandler, general counsel at the IT firm Cisco Systems in San Jose, California, told a forum at the American Enterprise Institute in Washington DC last month. "It's a huge increase in the amount of litigation and in expense."

Lying in wait

The majority of the cases he deals with today, Chandler says, are brought by 'patent trolls' who simply buy patents and then lie in wait for innocent companies to infringe them, adding no value or innovation themselves.

The proposed changes "bring patent law back to its focus of promoting innovation and not unfairly rewarding inventors at the expense of the economy", says Andy Cadel, the top intellectual-property lawyer at JPMorgan Chase, a financial-services firm based in New York.

The stark divide between those who make pills and those who make PCs brings considerable pressure to bear on senators with constituents in both camps, such as Joseph Biden (Democrat, Delaware) and Dianne Feinstein (Democrat, California). Those divided loyalties are just what the Biotechnology Industry Organization and its allies are counting on as they launch an all-out effort to block the bill from reaching a Senate vote. They face an uphill climb given the IT industry's funds and finesse — not to mention its friendly history with the Democrats who control Congress. ■

STRINGER INDIA/REUTERS

The little shrub that could — maybe

India, like many countries, has high hopes for jatropha as a biofuel source, but little is known about how to make it a successful crop. **Daemon Fairless** digs for the roots of a new enthusiasm.

With a top speed of about 110 kilometres an hour, India's Shatabdi Express is not much to brag about by the standards of a French TGV or a Japanese Shinkansen train. Nonetheless, as the stock for one of the country's fastest and most luxurious passenger lines, the Shatabdi trains have a certain prestige. So when, on New Year's Eve 2002, the Shatabdi train from New Delhi to Amritsar was powered in part with biodiesel for the first time, it was a clear statement of the government's desire to wean India off imported petroleum.

Diesel is India's main liquid fuel: the country burns roughly 44 million tonnes, or 320 million barrels, of the stuff a year, as opposed to about 94 million barrels of gasoline. The trains account for a significant part of that. Kunj Mittal, who heads the government-operated rail service's engineering and traction division, says its fleet of 4,000 engines currently burns about 1.7 million tonnes a year, and that he wants to replace at least 10% of that with biodiesel at some unspecified point in the future. But he would need 200 million litres of biodiesel a year. Which is a problem. "At this stage," says Mittal, "there is no mass production of biodiesel."

Like many others around India, the rail service is looking to an unprepossessing, poisonous scrub weed to try to do something about that. It has planted a million *Jatropha curcas* seedlings on unused land along its tracks and elsewhere. It's just one symptom of the jatropha fever that is spreading around the country and the world — to the slight bewilderment of some of the scientists who best understand the shrub.

Jatropha, a member of the euphorbia family, originated in Central America. It has long been used around the world as a source of lamp oil and soap, and also as a hedging plant. One of its great selling points as a biofuel is the fact that growing it need not compete with the cultivation of food. Of 306 million hectares of land considered in a report by India's Ministry of Rural Development, 173 million are already under cultivation but the rest is classified as either eroded farmland or non-arable wasteland. That's the sort of land that jatropha can thrive on, with bushes living up to 50 years, fruiting annually for more than 30 years and weathering droughts with aplomb[1]. In the early 2000s then-president A. P. J. Abdul Kalam repeatedly endorsed the plant for its potential contributions to energy security and as a route to greening barren land. Jatropha has been held to promise a reliable source of income for India's poor rural farmers and energy self-sufficiency for small communities — all while reducing fossil-fuel greenhouse-gas emissions and soil erosion.

In 2003, India's Planning Commission recommended a national mission on biofuel, a two-phase project for wide-spread cultivation of jatropha on wasteland across much of India. The first phase of the mission aims for 500,000 hectares of jatropha grown on government land across the country. The biodiesel would be produced primarily by panchayats — local governing bodies — at the village level, coordinated at the national level by a consortium of government departments. Should the first phase go according to plan, India's central government would embark on the second phase of the mission — planting a total of 12 million hectares of the plant and privatizing the production of jatropha biodiesel.

Although it seems likely to go ahead eventually, various ministerial meetings that might have given the national mission on biofuel the seal of approval have been postponed in favour of higher-priority issues. Despite this,

several states have enthusiastically hopped aboard the jatropha express, providing free plants to small-scale farmers, encouraging private investment in jatropha plantations and setting up biodiesel processing plants. The Ministry of Rural Development, which is set to coordinate the national mission on biofuel when it is approved, estimates that there are already between 500,000 and 600,000 hectares of jatropha growing across the country.

And India is not alone in its hopes for the shrub. In February 2007 China, which claims to have 2 million hectares of jatropha already under cultivation, announced plans to plant an additional 11 million hectares across its southern states by 2010. Neighbouring Myanmar (Burma) has plans to plant several million hectares; and the Philippines, as well as several African countries, have initiated large-scale plantations of their own. India looks forward to encouraging more such schemes and quite possibly profiting from them. "Once we have an operational programme and have something to offer the world," says Krishna Chopra, the recently retired principal adviser to India's Ministry of New and Renewable Energy, "I think exporting the know-how would certainly be one of the first areas to develop."

The great unknown

Although there is reason to be enthusiastic about jatropha's potential as a biodiesel feedstock in India and beyond, there is one rather sobering concern: despite the fact that jatropha grows abundantly in the wild, it has never really been domesticated. Its yield is not predictable; the conditions that best suit its growth are not well defined and the potential environmental impacts of large-scale cultivation are not understood at all. "Without understanding the basic agronomics, a premature push to cultivate jatropha could lead to very unproductive agriculture," says Pushpito Ghosh, who has been working on the plant for the best part of a decade, and who is now director of the Central Salt and Marine Chemicals Research Institute (CSMCRI) in Bhavnagar.

When Ghosh first arrived at the CSMCRI, the United Nations Development Programme (UNDP) had already given the institute funding for the cultivation of a modest jatropha plantation, although not for biofuels work. The idea was to see "how to make use of waste land, coastal areas and sand dunes", Ghosh says.

The plantation started off as an unirrigated, unfertilized, 20-hectare patch of exhausted scrub: Ghosh wasn't particularly impressed when he first saw it. "There were shrubs and they were growing," he recalls, "but it didn't look to me that it had what was required to make a successful plantation. 'Where are the seeds?' I said to myself. I didn't see too many of them. Merely planting and letting jatropha grow doesn't necessarily lead to productive growth." Nonetheless, the fact that jatropha lived up to its reputation as a shrub that could eke out a living on relatively barren land piqued the interest of India's Department of Biotechnology, which provided a little further funding for exploration of biofuel possibilities using cuttings from three of the most productive plants in the UNDP trial.

> **"I saw all this green in what is otherwise a complete desert."**
> — Klaus Becker

The seedlings were planted in small plots spread over patches of degraded, untended land in the eastern state of Orrisa. "The results were not outstanding," says Ghosh, "but they were consistent." Several plants yielded around 1.5 kilograms of seed, enough for about 0.4 litres of diesel. As modest as the results were, says Ghosh, they created a lot of interest. "For the first time," he says, "we were doing something in a systematic way."

The CSMCRI's work also caught the imagination of Klaus Becker, who arrived at the institute in 2000 as a visiting agricultural scientist from the University of Hohenheim in Germany. The original UNDP plot inspired him far more than it had the sanguine, measured Ghosh. "I saw all this green in what is otherwise a complete desert. There was absolutely nothing else around it. 'Look,' I told Ghosh, 'if you get this working, you'll be the first in the world'."

From seed to oil

Becker returned to Germany and set about fund-raising. By 2003 he had cobbled together a €1.7-million (US$1.9-million) research fund comprised of grants from DaimlerChrysler, the German Investment and Development Company in Cologne, India's Council of Scientific and Industrial Research and the University of Hohenheim. With these funds, Ghosh and his team — working in collaboration with Becker and scientists at DaimlerChrysler — began exploring the transesterification needed to turn jatropha into biodiesel. The process had already been established by Nicaraguan researchers during the 1990s[2] and it wasn't long before Ghosh and his team were producing small batches.

"You could tell simply by looking at it that it was fairly good quality," says Ghosh of their first attempts. Chemists at DaimlerChrysler's Stuttgart labs analysed it in more detail than the CSMCRI was able to and judged it easily good enough to meet European standards. Further tests at the Austrian Biofuels Institute (ABI), which pitted the CSMCRI's jatropha biodiesel against fuels from other feedstocks, showed that it "clearly outperformed biodiesel from rapeseed, sunflower and soya bean oil in [its lack of a propensity to oxidize]," says the ABI's Werner Körbitz, adding that the fuel "showed a fully satisfying performance concerning power, efficiency and emissions".

Ghosh's vision — and part of the CSMCRI's mandate — was to create a version of this transesterification process that was both inexpensive

J. CHIKARA

J. CHIKARA

Oasis in the desert: Jatropha cultivation can halt soil erosion, increase water storage in the soil and transform barren expanses into lush, productive land.

A CHOICE OF CROPS

Biodiesel crop	Litres of oil per hectare
Oil palm	2,400
Jatropha*	**1,300**
Rapeseed (canola)	1,100
Sunflower	690
Soya bean	400

Source: United Nations Development Programme/World Bank
*Indian Planning Commission

Jatropha is already under cultivation in Tamil Nadu, India, where it can be grown with other crops such as sunflowers.

G. JAWORSKY

and easily replicable at the village level. Nearly 80,000 of India's 600,000 villages currently have no access to fuel or electricity — in part because there isn't enough fuel for a fuel distribution network. "If people can grow oil directly in villages and produce biofuels themselves in decentralized plants," says Ghosh, "then they can achieve energy self-sufficiency. My colleagues and I are deeply committed to this principle."

"The constant urge to simplify and to ensure that every gram of jatropha is turned into something valuable was a tremendous motivator," he says, looking back at the project. But while he and his colleagues were still congratulating themselves on a job well done, the *Times of India* ran a story announcing that DaimlerChrysler was set to test two of its Mercedes C-Class cars on a 6,000-kilometre road test across the length and breadth of India using the CSMCRI's jatropha biodiesel.

Up the Khardungla pass

It was the first Ghosh had heard of it. "Our focus all along has been biodiesel as a fuel for village folk," he says, "not for fancy urban folk." And on top of that there was an obvious practical difficulty. Up to this point, Ghosh and his team had only ever produced a few litres of it at a time: you can't get across India on that.

> **"Our focus all along has been biodiesel as a fuel for village folk, not for fancy urban folk."**
> — Pushpito Ghosh

Within a few months, though, Ghosh's team had developed a transesterification unit capable of producing about 250 litres a day — adequate for use in villages and small-scale industry[3]. The Mercedes ran entirely on 100% jatropha biodiesel from this unit throughout April and May 2004 without any significant engine modifications. In the summer of 2005, DaimlerChrysler had several automotive journalists take the cars on a high-altitude test through the Himalayas, including Khardungla pass, which, at 5,359 metres above sea level, is one of the world's highest motorable roads.

While Ghosh and his colleagues were making sure that jatropha could be processed as a reliable source of biodiesel, several of India's state governments were busy promoting their own jatropha cultivation campaigns. The state of Chhattisgarh, which has the most well-developed biodiesel programme in the country, has distributed 380 million jatropha seedlings to farmers, free of charge, over the past 3 years, enough to cover 150,000 hectares with the shrub. Shailendra Shukla, executive director of the Chhattisgarh Biofuel Development Authority (CDBA), says the state has also provided 80 oil presses to various village panchayats, and guarantees to buy back jatropha seeds — which have to be hand-picked off the shrubs — at 6.5 rupees (about US$0.16) per kilogram in order to stimulate confidence in the crop. Several local businesses have popped up across the state, says Shukla, that are now operating micro-refineries. "These are small businesses that provide biodiesel for the use in tractors, irrigation pumps, jeeps and village power generators."

Ghosh says that the CSMCRI has received an order for a refinery from the country's Defence Research and Development Organisation, part of India's Ministry of Defence. He explains that the unit would be capable of producing about 1,000 litres a day and would cost about 14 million rupees to install. In such a plant, he says, each litre of biodiesel would have a net production cost of about 26 rupees if the seed pods are bought at 6 rupees per kilogram and every scrap of seed and seed pod is converted into something valuable, with the seed going into oil, the bi-product seed cake into fertilizer and the seed husk into a high-density brick that can be burnt for fuel.

The wide governmental support has also attracted substantial business interest. D1 Oils, a UK-based biodiesel producer, is the world's largest commercial jatropha cultivator, responsible for around 81,000 hectares of jatropha in Chhattisgarh and in the southern state of Tamil Nadu, with plans for an additional 350,000 hectares over the next few years. "The entire programme revolves around the government-funded jatropha seeds," says Sarju Singh, until recently managing director of D1 Oils India. "The government gives farmers free or subsidized seedlings and D1 Oils guarantees to purchase the seeds at the price prescribed by the state." The company claims to have invested more than £3 million (US$6 million) in plant science and financing its share of the plantings, which are joint ventures.

Cautious approach

Yet most of these plantings have yet to reach whatever maximum level of productivity they might eventually attain — the plants need a few years to bed in. And Ghosh is wary of subsidizing jatropha too much before mass cultivation of the plant is fully understood. "A lot of government funds may go down the tube," he warns. Ghosh doesn't want the farmers to take on too much risk, so he is suggesting that they intersperse jatropha between their current crops, rather than banking on it as a cash crop. Shukla has similar reservations. "My immediate concern," he says, "is that because the seeds are derived from wild plants there is no assurance of yield." Shukla says the CBDA, like Ghosh, is promoting jatropha as something farmers limit themselves to planting between their rice fields. The only situation where all are agreed that it makes sense for small farmers to cultivate whole fields of jatropha is on farm land that has become or is becoming unproductive. It is a good fallow crop, says Becker: "It has a deep root system which stops ground erosion and increases water storage in the soil." This, he says, leads in turn "to more biomass growth and an accumulation of organic carbon in the soil".

Henk Joos, D1 Oils' director of plant science and agronomy, agrees that assured yields and the techniques needed to achieve them on a

S. L. PUROHIT

large scale need a lot more research. Yield estimates currently vary a great deal. India's Planning Commission estimates about 1,300 litres of oil per hectare, but Ghosh, conservatively, foresees a figure of about half that. Yield research is the main focus of D1 Oils' Indian operations, he says. The company is currently testing a number of jatropha varieties to see which ones grow best in India's varied climatic zones. "It will be two or three years before we get real scientific data to base an industry on," he says. "We are not there yet, we have a lot of work to do."

Pushpito Ghosh tops up a vehicle that has covered 48,000 kilometres powered only by jatropha biodiesel.

This is the sort of work Ghosh is currently overseeing at the CSMCRI's test plots. "It isn't the most glamorous work, but the mass multiplication of reliably producing plants is key to developing an industry, he says. Ghosh and his team are looking at precisely what kind of soil conditions and just how much water jatropha needs in order to reliably pump out oil-bearing seeds. The fact that jatropha plants can survive droughts does not mean they will not be more productive if they get more water. The optimum amount of water is still unknown.

The team is also continuously on the lookout for plants that could be potential progenitors for a generation of a high-yield crop. "We have one plant which has given us 5 kilograms of seed," says Ghosh. "We have yet to get that from any other plant." The CSMCRI is trying to perfect the use of shoot-tip cuttings as a means for mass-replication of jatropha plants so it can capture their best attributes. Culturing tissue cuttings from the plant's growing tip, says Ghosh, is the most reliable means of propagating exact copies of a parent plant, an important step in creating an army of dependable high-yield clones. It's a common enough technique — but like so much technology, it hasn't yet been reliably adapted to jatropha. "The problem is, we just don't have the protocol right," says Ghosh.

These various efforts are not part of any overarching plan. Despite the general enthusiasm for India's national mission on biofuel, there is a definite lack of cohesion at the national level. "Right now, ad-hoc research is being done by different agencies," says Chopra, "but it doesn't add up, because they each do their own thing." A national biofuel policy that was written by Chopra and his colleagues shortly before his retirement might help. It envisages an authority that would coordinate research and provide funding through various government agencies in order to cultivate jatropha on an industrial scale. But this policy, like the national mission on biofuel, has yet to go through the cabinet. In this case, it has been stalled by disagreements between various ministries on how to price jatropha — the Ministry of New and Renewable Energy suggests subsidizing seeds; other government ministries suggest subsidizing biodiesel itself. But, says Chopra, "I expect it will come together, perhaps this year or early next year".

Ghosh remains cautious and optimistic in level-headedly equal measure. "We must neither get carried away by hype nor get despondent if the initial results of cultivation are not as per expectation," he says. "The future will depend on how seriously and scientifically we pursue our goals." ■

Daemon Fairless is this year's winner of the IDRC-Nature fellowship.

1. Francis, G., Edinger, R. & Becker, K. *Nat. Res. Forum* **29,** 12–24 (2005).
2. Foidl, N., Foidl, G., Sanchez, M., Mittelbach, M. & Hackel, S. *Bioresource Technol.* **58,** 77–82 (1996).
3. Ghosh, A. *et al. Int. J. Environ. Stud.* — special issue on India's future energy options (in the press).

See Editorial, page 637.

STRINGER INDIA/REUTERS

SCHOOL OF NUKES

How do nuclear inspectors know when all is not as they are told? **Geoff Brumfiel** joins some inspectors-in-training as they learn the ropes at the Los Alamos National Laboratory.

At half-past eight in the morning, the New Mexico sun already hot on our necks, we gather outside what could be the entrance of a high-security prison. Past the guard post and fence lies a tan, windowless building that covers about a city block. Gill-like vents along the building's side give it the look of a leviathan beached unnaturally in the high desert.

This is the Chemistry and Metallurgy Research building at Los Alamos National Laboratory, the United States' oldest nuclear-weapons laboratory. The 51,000-square-metre building is the weapons programme's main centre for the study of nuclear material.

But for the next two weeks the building will serve another purpose. It will be a training base for a fresh crop of nuclear inspectors for the International Atomic Energy Agency (IAEA). The IAEA is the United Nations' body charged with ensuring that the world's civilian nuclear facilities are being used for peaceful purposes, and it relies on about 250 highly trained individuals to do the job. The seventeen inspectors waiting at the gate are here as part of the agency's on-going training programme, honing the skills needed to stop the spread of nuclear weapons. The inspectors are "the eyes and ears on the ground", says David Albright, head of the Institute for Science and International Security, a non-proliferation group based in Washington DC. "And they're quite effective."

When I was younger, I entertained the thought that I, too, might want to be a nuclear inspector. I imagined travelling the world with a diplomatic passport, turning up unexpectedly at secretive facilities and matching wits with local despots. So when the IAEA offered to let me tag along on the first few days of this August's course I jumped at the chance.

My two days as a student there taught me that the reality is at once more tedious and more demanding than my Ian Fleming fantasy would have led me to believe. Inspectors are part scientist, part detective and part diplomat, says Jean Maurice Andre Crete, who heads the IAEA's training programme at their headquarters in Vienna, Austria. They also need to be part accountant, keeping meticulous logs of reactor operations, waste-pool inventories and research-material stocks. And they must be personable enough to win the trust of local plant managers. They must be Jacks and Jills of all trades.

> **"The inspectors are the eyes and ears on the ground."**
> — David Albright

To reach that goal, inspectors undergo a gruelling three-month induction and take continuing education courses in subjects as diverse as international law, psychology, satellite imagery analysis and environmental sampling throughout their career. But at the heart of it all is the need to be able to verify whether the canister in front of them contains the radioactive material that its custodians claim is inside. And that's why they train at Los Alamos. In the belly of the metallurgy behemoth, they have an opportunity to get to grips with pure weapon's-grade uranium and plutonium.

LANL

Blending in

If it wasn't for the fact that we're shuffling about in front of one of the world's most fortified scientific laboratories, we'd be an unassuming group. We come from 11 countries, spanning the globe from Argentina to Indonesia. All but two of us are men, and we're mostly middle-aged and dressed in polo shirts, khakis and tennis shoes.

Most are nuclear engineers, but a few are scientists. One of the younger inspectors, called Giuseppe, is an Italian radiochemist who was working at a national research institute near Milan before joining the agency last year. Another, named Valeriy, has a PhD in particle physics, although he has worked in the field of uranium enrichment for

decades, first in the Soviet Union and later in the Russian Federation.

Security is tight, so much so that we have to keep in our escort's line of sight at all times. We cross an immaculate front lawn distinctly out of place between the razor-wire and the formidable walls. A lone cooler sits on a picnic table. "That's the good thing about working behind the fence," jokes Peter Santi, one of our instructors. "Nobody's going to steal your lunch."

We wind down the staircase to our basement lab. Blast barriers line the hallways to protect heavily armed guards if attackers try and storm the building. But in contrast to the imposing exterior, the room in which the inspectors will spend the next two weeks is more like a college physics laboratory than a secure nuclear facility. The main room has eight tidy lab stations, each with its own instruments and worksheets. Friendly looking instructors chat casually over coffee. Only two things hint that something more dangerous is at play: a line of wall safes that hold canisters of high-purity uranium-235 and plutonium-239, and a plainly printed sign reminding users to be aware of 'criticality limits' — that is, the maximum amount of material that can be brought together without triggering a nuclear incident.

"A plainly printed sign reminds users to be aware of criticality limits."

Indirect comparisons

After introductions, David Bracken, head of training at Los Alamos, gives a brief overview of the next two weeks. The inspectors will learn about what is known in the business as a non-destructive assay, or NDA. An NDA is a suite of measurements that allows inspectors to determine both the quantity and composition of a material without ever sampling it directly. Done right, it can verify a country's stocks of nuclear material quickly and cheaply. "It's the foundation of safeguarding," Bracken tells us.

An NDA basically comes down to measuring two types of radiation — γ-rays and neutrons. The γ-rays, energetic photons emitted during nuclear decays, give a distinctive energy spectrum that is a fingerprint of a material's elemental composition. The neutrons provide a measurement of the quantity of material involved.

I haven't studied physics in nearly a decade, but I can follow the morning's lecture on neutron measurements without difficulty. It's a fairly simple technique: the material is placed at the centre of a doughnut of polyethylene foam. The polyethylene slows down the neutrons sputtering out of the sample so that they drift gently into a set of 18 or so cylindrical detectors placed around the foam. When the helium-3 atoms inside these detectors are struck by a neutron, they release a cascade of charged particles that goes on to be picked up by a high-voltage wire at the centre of each detector.

It's a low-tech, cheap and durable way of counting neutrons. Nothing fancy, nothing too complicated; reliable and portable. "The techniques aren't that new, and they're not that challenging," Santi says. "The challenge is to understand how they fit into the real world."

That afternoon, Santi's warning starts to make sense. A multitude of factors can affect the measurements. The distance of the detectors from the source; the thickness of the foam; the shape of the sample; the quality of calibration in the field: all these can trick an inspector into thinking that there is less, or more, or a different kind of material than is actually present.

Measuring γ-rays is also quite simple — we are given detectors just like the ones I used in my undergraduate lab work. But in the real world, shielding can block crucial features of a material's spectrum, and a dizzying array of common impurities, such as caesium and cobalt, can confound even the simplest detection. This is exactly why the inspectors are brought to Los Alamos, Bracken tells me. Only in a facility such as the metallurgy centre will the inspectors get the chance to work directly with realistically large quantities of weapon's-grade materials. "The way we teach here isn't just to push a button," Bracken says. "We try to teach the physics, teach them how to think."

Over lunch with Giuseppe, I learn that taking these measurements in the field is tougher still, because there's very little support. Only in the most confrontational cases, such as Iran and North Korea, do inspectors work in teams of the sort that appear on television. On visits to places such as Germany or Mexico they travel alone.

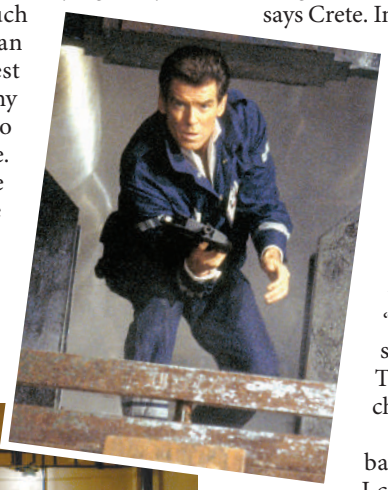
Field work

In nuclear power plants, especially, the inspectors must also work quickly. Commercial reactors are generally checked during refuelling outages, which companies try to complete as quickly as possible to maintain their profits. "You have just a few minutes," says Giuseppe. "And you have ten people standing behind you, talking in a language you don't understand."

Sometimes, the staff at a facility will try to restrict the inspectors' access, not necessarily to hide anything but simply to try and speed the inspection along. "If you're not prepared, people can sometimes prevent you from doing what you're trying to do," he says. "You have to be very sure about your rights."

Indeed, the rights of a nuclear inspector have changed radically over the past decade, says Crete. Inspectors used to have no authority to ask questions or search for additional, undeclared stocks. But all that changed in 1997, with the amendment of the Nuclear Non-proliferation Treaty — the international agreement under which the IAEA operates. "Now we are allowed to ask some questions," he says. They also have the right to check for hidden stock.

After my two days, I head back down from the mesa. I can see the attractions of the world my sort-of classmates are learning to negotiate. "It's a very dynamic kind of life," says the 37-year-old Giuseppe, happy with his career shift, with three months of every year on the road and a bag packed at all times for unexpected trips. "You don't know exactly what to expect," he says. But as someone who's both indiscreet and not much of a detail person, I wouldn't be the man for it. Better for me and non-proliferation both that I gave up the fantasy and took up the notepad. ■



DANJAQ/EON/UA/KOBAL COLLECTION

D. CALMA/IAEA

James Bond's world (top) is a long way from that of a nuclear inspector.

Geoff Brumfiel writes for *Nature* from London.

Plagiarism? No, we're just borrowing better English

SIR — The accusations made by arXiv that my colleagues and I have plagiarized the works of others, reported in your News story 'Turkish physicists face accusations of plagiarism' (*Nature* **449,** 8; 2007) are upsetting and unfair.

It's inappropriate to single out my colleagues and myself on this issue. For those of us whose mother tongue is not English, using beautiful sentences from other studies on the same subject in our introductions is not unusual. I imagine that if all articles from specialist fields of research were checked, similarities with other texts and papers would easily be found. In my case, I aimed to cite all the references from which I had sourced information, although I may have missed some of them.

Borrowing sentences in the part of a paper that simply helps to better introduce the problem should not be seen as plagiarism. Even if our introductions are not entirely original, our results are — and these are the most important part of any scientific paper.

In the current climate of 'publish or perish', we are under pressure to publish our findings along with an introduction that reads well enough for the paper to be published and read, so that our research will be noticed and inspire further work.

Ihsan Yilmaz
Physics Department, Çanakkale Onsekiz Mart University, Çanakkale, Turkey

Plagiarism: text-matching program offers an answer

SIR — The removal of almost 70 papers from the arXiv server on suspicion of plagiarism is dismaying (*Nature* **449,** 8; 2007). But, in a similar way to that currently being tested by the cooperative group of publishers CrossRef ('Academic accused of living on borrowed lines' *Nature* **448,** 632–633; 2007), the search technology that led to this removal could be used to reduce future problems.

Every paper submitted to arXiv could be examined by a search engine that looks for overlap or correlation with all previous arXiv submissions. If enough of a match is found, a message could be sent to the submitter, listing the work(s) in which similarities have been detected. Should the submitter wish to proceed with their submission, the program would notify the editorial board and trigger an automatic review. The submitter would also be given the chance to explain that the flagged papers were not copied or that the copying was for some reason legitimate.

Such a system would address the problem of plagiarism only among papers published in arXiv, but apparently that would already be an improvement. And although plagiarists might opt to copy and translate from foreign-language journals, or simply alter wording enough to pass muster, making it more difficult will at least discourage the lazier offenders.

As journals should welcome eliminating plagiarism at the preprint stage before publication, they could support the effort by giving the arXiv site search access to their own full-text databases.

John Bechhoefer
Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

Need to strike balance with industry–academia rules

SIR — It is appropriate to ask why universities permit alliances with industry, given that conflicts of interest can result. But your Editorial 'California dreaming' (*Nature* **448,** 388; 2007) and your News story 'California campuses resist industry restrictions' (*Nature* **448,** 394; 2007), questioning the commitment of the University of California's campuses to regulate conflicts of interest, overlook the efforts that the campuses are making to address the issue, as well as the benefits that some collaboration can bring to public health.

Under guidelines currently implemented at the University of California, Los Angeles (UCLA; see http://dgsom.healthsciences.ucla.edu/administration/guidelinesMain), all marketing materials, including free lunches and other gifts from the pharmaceutical industry, are banned. Industry representatives are allowed to visit only by appointment and even then, only outside patient-care areas.

Faculty at other campuses have proposed greater restrictions, including clamping down on the granting of money from industry to some faculty members. Although such restrictions would remove one source of bias, they would also eliminate the benefits to public health of some alliances with industry. Breakthrough treatments are increasingly the product of academic collaboration with the pharmaceutical industry. Industry usually takes the lead in developing new medications or devices, but when academic researchers have the opportunity to examine cutting-edge treatments, they can learn more about root causes of illness and lay the foundations for even more effective treatments.

We don't expect any one set of policies to immunize academic researchers from bias or conflicts of interest. What is needed is the kind of tripartite policy that we have already developed at UCLA: the elimination of practices (such as free lunches) that add little to the academic mission but may introduce conflicts of interest; the regulation of practices (such as research grants) that add to the academic mission but may be sources of conflicts of interest; and the education of faculty, staff and trainees on sources of bias and conflicts of interest in academic medicine, to enable them to maintain the highest ethical standards.

Andrew Leuchter, Gerald S. Levey
Office of the Dean, David Geffen School of Medicine, UCLA Center for the Health Sciences, Los Angeles, California 90095-7035, USA

Biosafety risk in health lab move to central London

SIR — The recent outbreak of foot-and-mouth virus around the Pirbright animal-health laboratory in Surrey should dissuade the UK Medical Research Council from pursuing plans to move the National Institute for Medical Research (NIMR) from its semi-rural campus to the centre of London.

The expertise and facilities of the NIMR, which are among the world's best, are vital to combat future infectious diseases, and are a key component of national and international biosecurity. Research carried out at the NIMR includes studies on emerging pandemic viruses, such as avian influenza (H5N1), which are grown in level-4 containment facilities.

The NIMR has a superb safety record. However, the recent crisis at Pirbright is a reminder that accidents happen. The current Mill Hill site is excellent for minimizing secondary risk. If the NIMR is moved to central London, the consequences of containment failure would be horrific. A three-kilometre exclusion zone, as used in the recent foot-and-mouth epidemic, would reach 10 Downing Street and quarantine most of the UK government's decision-makers. A ten-kilometre surveillance zone would affect all of central London.

Kings Cross, Euston and St Pancras railway stations, used by several million passengers each week, are five minutes' walk from the proposed NIMR site. After St Pancras becomes the London terminus for Eurostar in November, an escaped pathogen could reach Paris and Brussels within two hours. The first duty of government is security: the NIMR should stay at its present site.

Ellen Nisbet
Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 300 words; ideally shorter. Published contributions are edited. We welcome comments at Nautilus (http://blogs.nature.com/nautilus).

BOOKS & ARTS

An incomparable life

Exceptional intellect and creativity made Ernst Mayr the last century's greatest evolutionary biologist.

Ornithology, Evolution, and Philosphy: The Life and Science of Ernst Mayr 1904–2005
by Jürgen Haffer
Springer: 2007. 473 pp. $119

Jared Diamond

R. FRIEDMAN/CORBIS

Ernst Mayr at 98. At 100, he published his twenty-first book and the last seven of his 856 papers.

Returning from an expedition to New Guinea in 1965, John Terborgh and I laid out our hundreds of bird specimens in the Harvard Museum of Comparative Zoology for Ernst Mayr to identify. Ernst had made only one collecting trip to New Guinea 36 years previously, and his last publication on New Guinea birds had appeared in 1954. Nevertheless, as he walked along the shelf and glanced at one specimen after another, he quickly identified each by its Latin species name and then by its subspecies name; he told us which zoologist had described it, in what year and in which journal; gave the alternative names under which other zoologists had discussed it; and explained its broader biological significance (for example, "Check that one for altitudinal hybridization"). He hesitated only at one obscurely mottled specimen: "See if that's a female *Rhagologus*." We found later that it was indeed a female *Rhagologus*, a whistler whose relatives are usually banded black and gold.

This incident illustrates some of what made Ernst Mayr the greatest evolutionary biologist of the twentieth century. He is known especially for having woven together field studies of natural history, museum studies of taxonomy, and laboratory studies of population genetics to solve problems of the origin of species. He is also known for his syntheses of modern evolutionary biology and for contributions to understanding biology's distinctiveness within the history and philosophy of science. He released the last of his 21 books, *What Makes Biology Unique?*, on his hundredth birthday, after which he published the last seven of his 856 papers. He died five months short of turning 101. These achievements, plus his distinctive intellect and personality, make him an interesting subject for a biographer and historian of science.

Jürgen Haffer's is the first book-length biography of Ernst Mayr. Although others will surely follow, Haffer's will remain unique, as it was virtually co-authored by Mayr as a hybrid between an authorized biography and an autobiography. After meeting Mayr at the age of 64, Haffer interviewed and corresponded with him, and completed a draft in time for Mayr to review it. Haffer is himself a distinguished ornithologist and evolutionary biologist, so he understands and can explain Mayr's achievements. Like his subject, he is German-born but fluent in English, so he was able to read the young Mayr's diaries and letters in German and to appreciate how Mayr's German background and his American citizenship contributed to his science and his life.

A theme that emerges repeatedly from Haffer's account is the large role that chance played in Mayr's career — and the hard work and intrinsic ability that enabled him to profit from it. The first major stroke of luck was Mayr's field observation, at age 18, of a rare duck, the red-crested pochard. This led to his introduction to Germany's leading ornithologist, Erwin Stresemann, who helped the student Mayr to switch from medicine to biology. Mayr worked 18 hours a day to complete his PhD thesis in just 16 months.

Mayr had more good luck at his first lecture to a large general audience in December 1939. He was preceded by a dreadful one-hour lecture delivered by the geneticist Sewall Wright with his back to the microphone. Mayr's contrasting clear and splendid talk resulted in an invitation to write his first great book, *Systematics and the Origins of Species*. Students of creativity should compare Mayr with other scientists who also enjoyed good fortune and achieved distinction, but without equalling his greatness.

Haffer recognizes the qualities that contributed to Mayr's scientific achievements. Besides his outstanding memory for detail, instant visual pattern recognition, a synthetic ability to place details in a wider context, and legendary capacity for hard work, he was quick to learn (for example, he learned the Malay language from other passengers during his 26-day boat voyage from Italy to Java). He was determined in the face of resistance — as in his 11-year, ultimately successful, struggle to win US citizenship for his wife and himself from the anti-German New York immigration office. His social skills enabled him to survive in New Guinea, to befriend the tribes people there and amass rare bird collections from them, to establish hundreds of productive collaborations with other scientists, and to establish friendships with younger generations — such as Haffer and myself — as he outlived his own age cohort.

Mayr enjoyed a rare self-confidence. Haffer's book contains a long letter written by the 19-year-old Mayr to his mentor Stresemann, suggesting studies that Stresemann should pursue — quite shocking, considering the authoritarian posture of German professors to their students even today. Mayr confessed he was "terribly amused at my forwardness" when he re-read this letter 68 years later. The same self-confidence took him to New Guinea at age 23 to live as the only European in jungle camps with

New Guineans. This belief in himself rested on a realistic assessment of his own strengths and limitations, constraining him — unlike some other great scientists and many great musicians — to stay within his competence. For instance, he decided after many months of preliminary study not to extend his book *Animal Species and Evolution* to discuss plants, or *The Growth of Biological Thought* to include physiology and embryology, because he recognized his lack of familiarity with these subjects.

Mayr would defend his ideas vigorously. Often construed as dogmatism, it was a trait that he explained as heuristically useful by inviting challenge. He did not hesitate to discard his own long-held views when presented with convincing contrary evidence, as when the naturalist James Chapin persuaded him to abandon his original lamarckian outlook. Because Mayr had previously stressed the evolutionary significance of peripheral isolates, I expected trouble when, during our collaboration on our book *The Birds of Northern Melanesia*, I sent him an analysis demonstrating that such isolates rarely spread upstream in northern Melanesia — but he raised no objection.

This biography is required reading for evolutionary biologists, historians and philosophers of science, and for scholars of creativity. It is organized for such a readership: Haffer separates the phases of Mayr's career and the fields of his science, explains the scientific problems that Mayr studied, and appends Mayr's complete *curriculum vitae*, bibliography and an analysis of his publications.

Room remains for another biography of Mayr aimed at a broad public interested in how creative and productive minds are formed. Much more can be extracted from Mayr's diaries and letters, and from reminiscences of his students and younger friends. My advice to such a biographer: start soon, while those people are still available — and be grateful that Haffer has done so much of the groundwork in this splendid account. ■

Jared Diamond is professor in the Department of Geography, University of California, Los Angeles, California 90095-1524, USA.

The dark side of cancer research

The Secret History of the War on Cancer
by Devra Davis
Basic Books: 2007. 304 pp. $27.95

Daniel S. Greenberg

Scorn and bitterness steam from the pages of *The Secret History of the War on Cancer*. This amalgam of history, speculation and memoir argues that "the wrong battles with the wrong weapons and the wrong leaders" have consigned millions to preventable death from cancer.

Dangerous carcinogens — mainly tobacco, radiation, asbestos and benzene — continue to pollute the environment decades after their lethality was clearly identified, writes Devra Davis, an epidemiologist and director of the Center for Environmental Oncology at the University of Pittsburgh Cancer Institute in Pennsylvania. Their prevalence, she asserts, is the product of corporate greed and guile, warped political priorities, supine regulatory practices and dirty dealing by notable scientific generals. Today, new hazards may lurk in innumerable chemical compounds in convenience items, electromagnetic propagation and industrial waste, all shielded from scrutiny by malign cover-ups and evasions.

The well-known reality of this 'war' is that environmental clean-up, which means friction with corporate powers, has been neglected relative to curative strategies, which raise hope, enrich the scientific enterprise and offend no one. But the imbalance has been diminishing. In Davis's telling, however, little has changed in decades, leaving industrial polluters unrestrained, while devious characters profit by spreading cancer. For a nuanced understanding of the confrontation with cancer, look elsewhere.

Davis notes that Richard Doll, the epidemiologist credited with establishing the relationship between tobacco and lung cancer, held lucrative consulting deals with major chemical companies and industrial associations, including a firm that he "defended … against lawsuits from some of its asbestos-exposed workforce". By way of contrast, Davis chronicles the fate of several scientists who sounded alarms about environmental carcinogens, risking and sometimes losing their careers, while their findings were flushed down the memory hole.

Turning to her own experience, the author reports a conversation that she says occurred in 1986 — 22 years after the US Surgeon General's historic report on smoking and health, and 15 years after President Richard Nixon and the Congress declared war on cancer. While she was an environmental staffer at the National Academy of Sciences (NAS), Davis alleges that the then NAS president Frank Press doused her plan to write a book about "the fundamental misdirection of the war on cancer". He warned her, she writes: "You can't write a book critical of the cancer enterprise and hold a senior position at this institution." Davis remained at the academy for a decade. "I watched … the concerted and well-funded effort to identify, magnify, and exaggerate doubts about what we could say that we know [about carcinogens] as a way of delaying actions."

Asked to comment, Press, now an official at a consulting firm in Washington, wrote to me: "I don't recall the incident. It could have happened." Press added, "If as a staff member she wrote a book on issues before the academy, the NAS would be viewed as biased and predictable. Davis should have known this."

Davis details what was long-ago recognized about major environmental cancer risks, and how the polluting miscreants eluded control. The tales she tells, and retells, are generally well known, as evidenced by the predominance of published sources in 22 pages of citations. From insider whistle-blowers and corporate documents unearthed in legal proceedings, the monumental deceits of the tobacco industry, reiterated at length here, have been described and dissected in several distinguished books and innumerable articles. In this respect and others, the 'secret' in the title is questionable.

Nonetheless, for a well-documented, prosecutorial account of the dark side of cancer-control politics, Davis's work — lopsided and verbose as it is — merits attention. Younger readers, particularly, may be unaware of the corporate and political machinations that kept carcinogenic pollutants uncontrolled long after their dangers were understood.

Presented in fascinating detail is the long and troubled career of Wilhelm Hueper. Between the First and Second World Wars, this German émigré pathologist pioneered some of the earliest identifications of industrial

Smoking risks were suspected when this ad came out in 1946.

carcinogens. Hueper was threatened and eventually fired over his investigations of worker exposure at the chemical company DuPont. At the US National Cancer Institute (NCI), Hueper pursued the connection between environment and cancer, before meeting industrial resistance. Drawing on Hueper's unpublished autobiography, Davis relates that industrial firms were "given extraordinary access to his papers prior to their being submitted for publication when he worked at NCI". Permission to publish was denied, and Hueper was directed to confine himself to animal studies.

Pessimistic about industry voluntarily mending its carcinogenic abuses or government forcing a cleanup, Davis proposes an intriguing remedy. The creation of a Truth and Reconciliation Commission (TRC), à la post-apartheid South Africa, would, she reckons, encourage industrial executives to confess to their environmental delinquencies: "The files of many large multinational businesses could easily tell us about many more health risks associated with workplace exposures of the past." Under a TRC regime, she suggests, "grace and forgiveness become the grounds for renewal and restoration".

Davis's inventory of long-standing corporate and government tolerance of known carcinogenic exposures is faultless. Her failure to acknowledge that in important respects the tide is turning detracts from the credibility of her work. Her call for a TRC is a strikingly imaginative gambit. Given the power and mood of corporate America, file it under 'fantasy'. ■

Daniel S. Greenberg, a journalist in Washington, is author of *Science for Sale: The Perils, Rewards, and Delusions of Campus Capitalism*.

Promoting capitalism over Kyoto

Break Through: From the Death of Environmentalism to the Politics of Possibility
by Michael Shellenberger and Ted Nordhaus
Houghton Mifflin: 2007. 256 pp. $25

Jonathon Porritt

In 2005, Ted Nordhaus and Michael Shellenberger caused a furore among environmentalists when they published an article entitled 'The Death of Environmentalism' in the online magazine *Grist*. It combined penetrating insights into the inadequacies of contemporary environmental thinking with a robust, upbeat assessment of the opportunity to transform the global economy. At the time, it was exactly what was needed. But in elaborating their original thesis into a densely argued book, the self-styled 'bad boys of environmentalism' have largely lost the plot.

Their analysis of the historical relationship between affluence, economic security, personal identity and 'the American dream' is illuminating, particularly perhaps for European readers. Some of the early chapters assert a creative world view — on how best to deal with the rainforests in Brazil, for instance, or on the contrast between faith communities and the environmental movement. And their sustained, if repetitive, critique of conventional environmental thinking is still apposite. Many environmentalists are indeed too misanthropic, too focused on technocratic and regulatory solutions, and far too apocalyptic for their own good, let alone the good of the environment.

But Nordhaus and Shellenberger did all that three years ago. In *Break Through*, they overegg the pudding with grotesque exaggerations and generalizations, and with such a startling lack of generosity as to make one question their true motives. For instance, they would have readers believe that they alone are advancing an investment-led approach to climate change (rather than the limits-based regulations embodied in the Kyoto Protocol). Not a mention of *Natural Capitalism*, the radiantly upbeat book by Paul Hawken, Amory Lovins and L. Hunter Lovins, or of the 30-years-plus of inspired, investment-driven technology developments promoted by the Rocky Mountain Institute and Amory Lovins. Similarly, it would have been proper to commend the sophisticated global Marshall plan that Al Gore outlined in his 1992 book *Earth in the Balance* and to highlight the failure of Gore's documentary film *An Inconvenient Truth* to talk solutions.

The main disappointment is that *Break Through* does not produce any coherent new environmental agenda. There are a few references to the Apollo Project on clean energy, an excellent US initiative led by Nordhaus and Shellenberger, which unfortunately failed to gain political traction. Much more time is spent demolishing the environmental justice movement and discussing the inadequacies of 'not in my back yard' politicians than in spelling out how the environmental dream might actually become reality.

Also galling is the almost limitless flow of optimistic, unsubstantiated assertions about the benefits of a positive mindset. The authors urge environmentalists, instead of being so anti-growth, to proselytize that "things need to get a lot better economically before they can get better environmentally". I happen to agree with this view in terms of emerging countries such as China and India, let alone some of today's poorest nations, but Nordhaus and Shellenberger seem to position themselves outside the critical debate about the kind of economic growth required.

The political naïveté can be as painful as the spurious optimism. The damage done by seven years of the Bush administration is barely referenced. Capitalism is simply assumed to be a good thing, without any deeper analysis. Crude either/or dichotomies rule. For instance, anyone who sees the Kyoto Protocol as part of the global effort now required to decarbonize our economies is portrayed as dumb. Their idea that Kyoto's limits-based, target-driven process is somehow incompatible with investment in clean technology is the kind of simplistic illogicality that Tony Blair struggled with over many years' dealings with Bush.

Break Through is an important but massively flawed contribution to today's debate about the future of the environment. Its historical analysis of the movement is valuable. The promised radical alternative remains elusive — and would still be of enormous worth. ■

Jonathon Porritt is founder director of Forum for the Future, chairman of the UK Sustainable Development Commission and author of *Capitalism as if the World Matters*.

S. WERMUTH/REUTERS

To highlight climate change, naked volunteers pose for a photo on a Swiss glacier as part of a campaign.

Academic snakes and ladders

The Chicago Guide to Landing a Job in Academic Biology
by C. Ray Chandler, Lorne M. Wolfe and Daniel E. L. Promislow
University of Chicago Press: 2007. 176 pp. $14

Pamela A. Silver

At one time, academic biology was the province of those willing to work long, irregular hours and to wear threadbare jackets with elbow patches. Overachievers with an interest in biology generally went into medicine. Academic jobs were few, but demand was low. With the advent of biotechnology, the sequencing of the human genome, and overhead charges on government grants, faculty positions in biology have now transformed from a calling into a well-paid career. It was perhaps inevitable that 'how to' books would materialize to guide hungry job-seekers in what has become a competitive market.

The Chicago Guide to Landing a Job in Academic Biology is a nuts-and-bolts book, full of practical advice such as how to construct an application that won't be thrown out, how to give a seminar, and how to complete the interview process successfully. Some will profit from its advice and will be glad that they read the book. Others may be disturbed by its implicit message that landing a job in academia requires a careerist mindset and not just raw brilliance.

The book has a limited target audience — the fraction of graduate students and postdoctoral fellows that actually want to carry on in academic research. Starting with the assumption that the reader wants an academic job, the authors go into great detail about how to get one. What should you put in your cover letter with your job application? What should you wear at your interview? How many jobs should you apply for? The advice is sensible, being based on years of reading flawed job applications and watching candidates make fatal errors.

For someone who has already been through the process, it is hard to read the book without pondering the bigger questions. Is this really the best process for choosing faculty members? Does it select for the best researchers and teachers, or does it generate careerists who are good at applying for jobs? Does it favour advisers who are willing to write exaggerated letters of recommendation, thereby gaining a reputation as a good career-promoter? In a training process that may take six years of graduate school and four years as a postdoctoral fellow, whose interests are really being served?

As the authors point out, a biologist undergoes a lengthy apprenticeship that rarely includes advice on training for a job *per se*. Speculating why this should be so is beyond the scope of the book. The answer may be that for an adviser to discuss the issue in any detail is like a parent talking to a child about their mid-life crisis. The job application process can itself be embarrassing — many applicants are unsuccessful and the present system breeds conservatism. More to the point, any realistic discussion on the topic should also consider unfairness, bad luck, poor choices and compromise. A book that simply describes the mechanics puts the onus on the next generation to refocus the process in a way that nurtures excitement about the unanswered questions in the biological sciences. ■

Pamela A. Silver is professor of systems biology at Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

EVENT

Experiment-a-thon

Matt Brown

A giant sombrero lurks in the undergrowth of London's Kensington Gardens. Inside, a conical auditorium resonates with strange noises and good vibrations. The Serpentine Gallery's summer pavilion is gearing up for a 24-hour 'experiment-a-thon'.

The pavilion is an annual tradition. Leading architects with no previous commissions in the United Kingdom take a turn at designing the temporary venue. Alumni include Zaha Hadid (2000), Daniel Libeskind (2001) and Oscar Niemeyer (2003).

This year's steel and plywood structure is the work of Norwegian architect Kjetil Thorsen and Danish–Icelandic artist Ólafur Elíasson, whose installation 'The Weather Project' filled the Turbine Hall of London's Tate Modern gallery in 2003. The duo's pavilion has a spiral ramp around the outside, leading to a peaked roof designed to amplify sound.

The Serpentine Gallery is renowned for filling its edgy summer structure with original events. This season, the emphasis has been on 'experiments' — in which scientists, artists and architects exchange ideas from their fields with the public in Friday night 'laboratories'. These have included the debut performance of a new type of electric violin, and the pavilion itself played as an instrument. They are warm-up acts for the forthcoming all-nighter.

This weekend, around 30 performers, including autism researcher Simon Baron-Cohen and musician Brian Eno, will entertain and educate beneath the pavilion's strange geometries from midday on Saturday to midday on Sunday.

L. HAYES

Those planning to sit through the whole night can rest assured: the coffee is excellent. A similar event in Berlin follows next month.

Olaf Blanke of the Swiss Federal Institute of Technology in Lausanne will demonstrate his recent headline-making video technology for simulating out-of-body experiences. Other experiments deal with phantom limbs, colour vision and Capgras syndrome — the belief that close acquaintances have been replaced by impostors. Perception is an underlying theme. "Each experiment explores how the brain simplifies stimuli into a stable impression," explains participant Israel Rosenfield from the City University of New York.

"We're building on the success of last year's 24-hour interview marathon," says Hans Ulrich Obrist, co-director of exhibitions and programmes at the Serpentine Gallery. In Rem Koolhaas's 2006 pavilion, Obrist conducted a series of conversations with prominent Londoners. "A certain audience attends science events, another attends arts events; we hope to attract both." ■

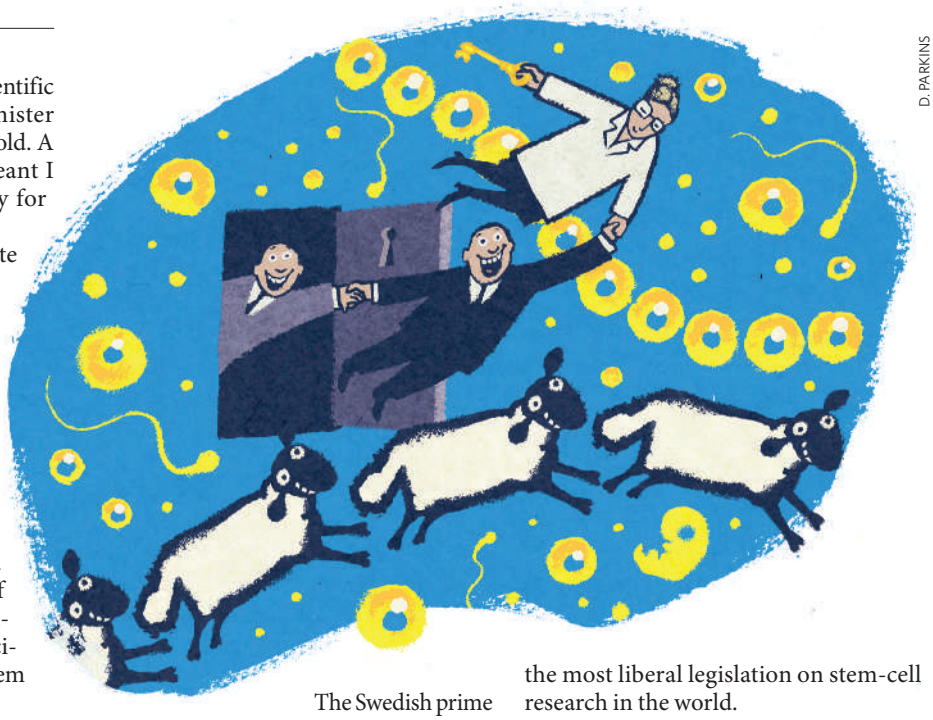
Matt Brown is the editor of *Nature Network London* http://network.nature.com/london.

The Experiment Marathon begins at noon on 13 October. A second part takes place in Berlin on 9 November.

When ministers are well primed

Enthusing and informing government members about science can have surprising and gratifying results.

Hans Wigzell



D. PARKINS

In 1999, I was appointed special scientific adviser to the Swedish prime minister and government, a position I still hold. A fortuitous lack of instructions meant I largely designed my own strategy for supporting the administration.

My aim was to educate and excite government members about science. The various schemes I introduced included ministers meeting with Nobel laureates and Sweden's best scientists, and retreats for the prime minister and minister of education with 10 selected researchers at a time. In these ways, I delivered more than 100 hours of top-quality science education to the prime minister between 1999 and 2006 and shared with members of the government my personal knowledge of some of the best Swedish scientists and science anecdotes for them to recount when needed.

During my time as special adviser I have learnt that day-to-day politics is often based on chance. The immediate availability of a trusted scientific adviser is essential in helping politicians to make decisions. When news broke about the cloning of Dolly the sheep, for example, I received an agitated phone call from Sweden's minister of education. He expressed his horror and was keen to respond rapidly to forbid human cloning. I asked him whether he had met any cloned people, to which he replied, "certainly not". Once I had explained that monozygotic twins are 'clones', he calmed down; no rushed announcement was made nor legislation introduced.

Of particular significance to Sweden was President George W. Bush's announcement on 9 August 2001 that only work on existing human embryonic stem-cell lines would be granted US federal funding. The National Institutes of Health helped the US government by declaring that more than 60 such cell lines were available globally. Roughly one-third of these lines were in Swedish laboratories — partly because Sweden had been a pioneer in developing *in vitro* fertilization technology. The Swedish media became excited: here was an opportunity for Sweden to produce medical research of great benefit to the country and mankind.

"Asked what a stem-cell line was, one minister said it looked like a Christmas tree with blunted twigs."

The Swedish prime minister asked members of his government whether they knew what a stem-cell line actually was. One thought it was cells growing in a straight line, a second described a blob of cells and a third said it looked like a Christmas tree with blunted twigs. So the prime minister asked me to lecture the government for 20 minutes on human embryonic stem cells during their Thursday session. The presentation created quite a stir and stimulated many questions; I left an hour and a half later.

The events that followed verged on surreal. The minister of defence flew north and included embryonic stem cells in his talk to the military. Two weeks later, the minister of industry rushed up to me and relayed the one high point of an otherwise uneventful meeting in Brussels. A minister had leaned over to him during lunch and boasted that his country housed two human embryonic stem-cell lines; his Swedish counterpart had proudly retorted, "we have 21".

The week after the Thursday session, Sweden's leading daily newspaper featured a full-page debate in which the ministers for education and health jointly proposed a permissive law for Sweden on stem-cell research. Eventually, this resulted in our country having some of the most liberal legislation on stem-cell research in the world.

The mix of schemes that evolved from my original mission to enthuse and inform ministers about science has so far taken many an unexpected turn.

The Swedish government, while chairing the European Union in 2001, had an important meeting in Uppsala with the EU research ministers. It was agreed that some cultural event should also be included. I had been acting at Stockholm City Theatre since 1989 and had appeared — with three of Sweden's best actors — in a play concerning misconceptions about molecular biology. So the government asked me to stage a version of the piece, starring all the ministers.

The Swedish government was thus, in different ways, inspired to promote the founding of a European Research Council (ERC). It continued to support this proposal avidly through to the final 2005 agreement of the European Union.

It is amusing to think that George W. Bush may inadvertently have supported the creation of the most progressive law so far on human stem-cell research and assisted in forming the ERC. But that's politics for you. ■

Hans Wigzell is senior strategic adviser to the Karolinska Institute, 17177 Stockholm, Sweden, and adjunct professor at Ehime University in Japan and the University of Baltimore, USA.

For more essays and information see http://nature.com/nature/focus/scipol/index.html.

Fifty years of hopes and fears

Atomic energy was cutting edge when the Windscale fire showed the world the effects of a nuclear accident. Fifty years on, we have more innovative ways to generate electricity.

Walt Patterson

On the edge of the Lake District, scenic heart of northwestern England, lies a vast, ugly industrial scar. Part is Calder Hall. The rest was once Windscale. Half a century ago both became famous worldwide.

On 17 October 1956, Queen Elizabeth II switched on the electricity from Calder Hall, called the world's first nuclear power station. On 8 October 1957, the physicist in charge of the Windscale Number 1 plutonium production reactor, a few hundred metres from Calder Hall, threw a switch too soon. The reactor caught fire. By the time the fire was extinguished three days later, a plume of radioactivity had drifted far across northern Europe. The world found out what a nuclear accident could do.

Last month, wreckers demolished the cooling towers of Calder Hall, on the site now known as Sellafield. The hopes — and fears — that those towers came to represent still loom over global energy.

In the 1950s, politicians, media and the public were keen on 'atomic power', as they called it. They felt it was cutting edge, a cleaner, more modern replacement for coal. Those keeping the lights on were less keen. In the United States, Britain and elsewhere, electricity managers viewed nuclear power, untried and unfamiliar, with scepticism. Their wariness deepened after the Windscale fire left a radioactive mess, still being cleaned up half a century later.

Nuclear promoters, backed by their governments, prevailed. Electricity demand and supply expanded dramatically. Nuclear plants, scaling up at breakneck speed, reached record size. They also had record cost and timetable overruns, sometimes tripling initial estimates and taking more than a decade to build. A brief flurry of orders followed the first oil crisis in 1973, then petered out, amid many cancellations. The last US plant completed was ordered in 1974.

On 28 March 1979, Three Mile Island 2 in Pennsylvania came scarily close to meltdown — the first major accident at a civil nuclear power plant. Soon nuclear order books were almost empty. On 26 April 1986, the Chernobyl 4 reactor exploded, the worst nuclear accident in history. By then, however, nuclear power was already falling out of favour, not for the oft-cited reasons of safety or waste management, but because of its cost and complexity.

SELLAFIELD

The cooling towers of Calder Hall at Sellafield being demolished last month.

For three decades after Calder Hall, nuclear power plants fitted the traditional electricity system, in which a better power plant was always a bigger one farther away. Then, at the end of the 1980s, governments began selling electricity assets to private operators and introducing competition. That transferred risk from the captive customers of traditional electricity monopolies to the shareholders and bankers of the new competitive systems. Private investors lost interest in nuclear plants. Instead they bought gas-turbine power stations that could be ordered, built, operating and earning revenue in less than two years. Suddenly a better power station might be a smaller power station closer to users — a fundamental break with tradition.

From 1990 onwards, gas-turbine generation has taken the lead. It is cheap, clean and easy to site. It needs no fuel store, uses less water than coal-fired or nuclear generation, and produces no waste. Gas turbines have also paved the way for other smaller-scale generators. Traditional water and steam power kept scaling up because a bigger unit made cheaper electricity. For microturbines, gas engines, Stirling engines, flow batteries, fuel cells, wind turbines, microhydro and marine energy, biomass power, solar thermal and photovoltaics, what counts is scaling up the number of units: the more you make, the cheaper their output, with rapid learning curves. Some also deliver both electricity and heat from the same fuel, boosting efficiency and reducing emissions. Generation from wind, water and sunlight uses no fuel and produces no emissions.

Traditional electricity networks are radial and one-way. They deliver large flows of electricity over long distances from huge power plants in remote locations. Such networks are inherently vulnerable to disruption, as widespread blackouts attest. Smaller-scale generation closer to users prefers two-way networks, linking loads and generation in optimized local systems. Such decentralized electricity offers higher performance and reliability, and lower environmental impact. It is gaining ground rapidly.

As climate and fuel security dominate the energy agenda, the battle between traditional and innovative electricity intensifies around the world, notably in fast-growing economies such as China. After half a century, nuclear power is the ultimate in tradition. It needs climate more than climate needs it. To avert catastrophic global warming, why pick the slowest, most expensive, most limited, most inflexible and riskiest option? In 1957, despite the Windscale fire, nuclear power was worth trying. We tried it: its weakness proved to be economics, not safety. Now nuclear generation is just an impediment to sustainable electricity. ■

Walt Patterson is associate fellow at Chatham House, London SW1Y 4LE, UK and author of *Nuclear Power* (Penguin), available free at www.waltpatterson.org. His new book is *Keeping The Lights On: Towards Sustainable Electricity*.

ESSAY

NEWS & VIEWS

LINGUISTICS

An invisible hand

W. Tecumseh Fitch

Quantitative relationships between how frequently a word is used and how rapidly it changes over time raise intriguing questions about the way individual behaviours determine large-scale linguistic and cultural change.

In the eighteenth and nineteenth centuries, linguistics was considered a thoroughly historical science, focusing on how languages such as English or Sanskrit changed through time. By uncovering rules governing phonological change, historical linguists reconstructed dead protolanguages such as Indo-European — an ancestral dialect spoken some 10,000 years ago that diverged into a wide variety of modern languages, including Hindi, Russian, Spanish, English and Gaelic. The crowning achievement of these early linguists was a family tree of languages that became an inspiration for Charles Darwin as he pondered biological evolution (Fig. 1). Unfortunately, many historical linguists entertained quasi-mystical ideas: August Schleicher, the German author of the first great comparative grammar of Indo-European languages, believed that languages are living things, and Jacob Grimm posited a *Sprachgeist* — an internal spirit of a language driving it to change along certain lines.

Twentieth-century linguists rejected such fanciful notions, and emphasized the capacity of individuals to produce and understand utterances. Noam Chomsky famously characterized this as a conceptual shift from a historical preoccupation with 'E-language' (a set of externalized utterances) to an emphasis on 'I-language' (principles internalized by the language learner)[1]. The focus by modern linguists on the individual capacity to acquire and use language (ontogeny) led to a flowering of research, allowing the biological and neural underpinnings of language to be studied coherently, and opening the door to consideration of how the language faculty evolved biologically (phylogeny). But this approach left behind the traditional questions of the cultural evolution of individual languages (glossogeny[2]) that tantalized earlier generations of linguists.

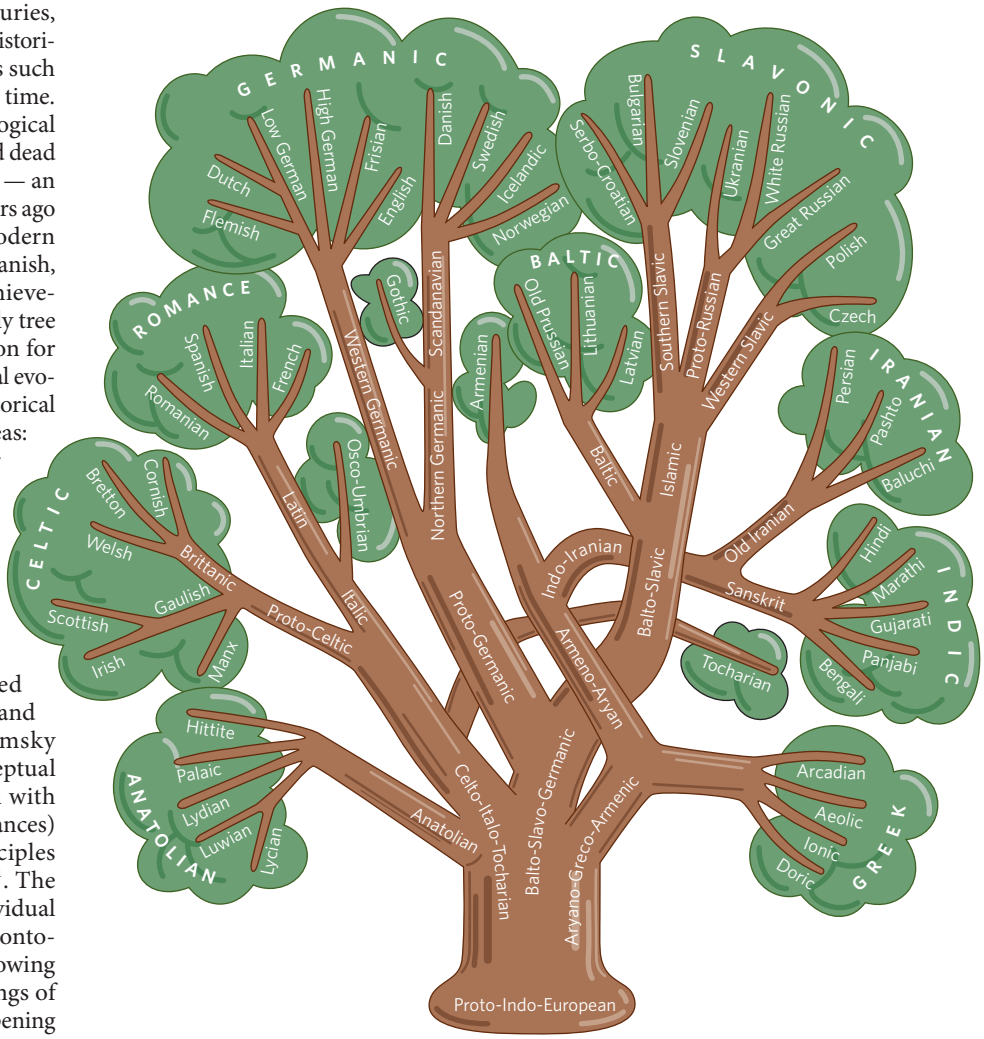


Figure 1 | The glossogenetic tree of Indo-European language. The words of language are not inherited biologically, but are passed on culturally through learning. This process of 'cultural evolution' generates a hierarchical tree of relationships among languages, here illustrated by the Indo-European family. Just as descent with modification in biological evolution (phylogeny) leads to phylogenetic trees, so the analogous process in language change (glossogeny) can lead to glossogenetic trees.

Elsewhere in this issue, two papers[3,4] revisit these issues from a fresh perspective. Both concern language change, and come from laboratories of well-established evolutionary theorists. Both analyse historical linguistic data to show that patterns of change depend strongly on the frequency with which words are used in discourse, as measured from large contemporary databases. Lieberman *et al.* (page 713)[3] consider the cultural evolution of the English past-tense marker '-ed'. In Old English, this was just one of many different rules used to indicate times gone by. Today, the other once-widespread rules remain only as irregular residues, such as 'fly/flew/flown'. By tracing their disappearance, the authors derive an exact quantitative relationship between the frequency of verb use and the speed of this pruning process: a verb used 100 times more often than

Box 1 | The invisible hand in language change

Language change at the 'macroscopic' level is often influenced in counter-intuitive ways by 'microscopic' changes in how individuals use language. A nice example is found in the historical phenomenon of pejoration in words referring to women, where respectable words acquire negative connotations over the centuries. A 'hussy' was once a perfectly respectable housewife, and 'wench' just meant 'young woman', but both terms now connote a woman of loose morals. And 'lady' — once used just for a woman of noble birth — is now the standard term for any woman.

Intriguingly, words for men generally don't suffer the same fate, and sometimes even improve their connotations ('knight' originally meant just a boy or retainer). Parallel patterns have occurred in other languages (for example, as with the German *Weib*, which suffered the fate of 'wench').

The most obvious explanation for this phenomenon is that language users (or at least those who have historically been responsible for recording language — men) are consistently misogynistic. But a more convincing 'invisible hand' explanation invokes a simple individual rule: when talking to or about women, err on the side of politeness[8]. Given two options, one normal and one polite ('hussy' versus 'lady'), this rule, if applied widely and consistently, leads to 'lady' becoming the common form. 'Hussy' or 'wench', by comparison, become ever-less polite over time. The best intentions lead to pejoration as an unintended consequence. **W.T.F.**

another will regularize 10 times more slowly.

Pagel *et al.* (page 717)[4] take a broader approach, quantifying the rate at which related words (such as 'water' in English and *Wasser* in German) have been replaced by other forms (such as the French *eau*) during the cultural evolution of 87 Indo-European languages. Using frequency data from four different language corpora — sets of texts representing patterns of usage in English, Spanish, Russian and Greek — and sophisticated tree-based statistical methods over the whole glossogenetic tree, Pagel's group derives a relationship holding over millennia. The relationship explains 50% of the variation in replacement rates between different words — a level of statistical power rarely observed in the social sciences, particularly across a wide range of cultures.

Despite significant differences in their methods, both papers document the same general pattern: frequently used words are resistant to change. Relatively infrequent inflections such as 'help/holp' became regularized, whereas high-frequency English verbs retained their ancestral irregular state ('go/went' or 'be/was'). More generally, terms that occur with high frequency in Indo-European languages (such as 'one', 'night' or 'tongue') are resistant to substitution by new phonological forms. The realization that frequency of use has a significant role in language change is nothing new[5–7]. But the use of sophisticated methods developed in bioinformatics and genomics to quantify these relationships is an important step forward. We can expect similar approaches to be applied to a wide variety of languages, to determine whether the specific patterns uncovered in these papers also hold in non-Indo-European languages, such as Chinese or the Dravidian languages of southern India.

Documenting these relationships remains descriptive, not explanatory: quantifying their form does not tell us why such regularities exist. Schleicher might have characterized the situation with an E-linguistic metaphor, in terms of a struggle for survival among different word forms. But from an I-linguistic perspective, Chomsky would retort that the underlying explanation must come down to the individuals who learn and use the language. Pagel and his colleagues[4] consider two such possibilities. First, new phonological forms might arise less often for high-frequency words because errors of perception, recall or production are less common for frequently used words. Alternatively, such cultural 'mutations' might occur uniformly, but frequency of use would affect the probability of new variants being adopted by the population.

Crucially, these two possibilities are not necessarily in conflict. An adequate explanation for glossogenetic phenomena must incorporate individual and collective levels of description, and show why they are necessarily related. Part of the challenge is the apparent circularity of explanation inherent in evolutionary systems, where the output of one generation serves as the input to the next, and causes are no longer neatly separated from effects. Such difficulties are compounded in cultural evolution: glossogeny represents an intermediate descriptive level that changes on a slower timescale than ontogeny, but faster than phylogeny. Although we, as individuals, don't generally invent words or grammatical forms, our usage (mispronunciations, or slight semantic shifts, for example) will affect their future transmission, and the population's usage en masse will determine their fate across many generations. Thus, human languages such as French or Swahili are neither natural (as envisioned by Schleicher) nor artefacts made intentionally by individual humans. Like economic, political and religious systems, they are phenomena of a third kind[8].

Although this distinction is intuitive (we apply it when distinguishing 'natural' languages such as English from 'artificial' ones such as Esperanto or C++), it remains relatively unexplored in linguistics. Theories concerning such phenomena have been developed more extensively in economics: the necessity of explaining 'macroscopic' phenomena in terms of quite different 'microscopic' behaviours was first discerned by Adam Smith, who used the evocative metaphor of the "invisible hand" to describe how individuals working to maximize profit in their own self-interest benefit society as a whole by driving up the standards of goods and services on offer.

Where should we look to gain a deeper understanding of the invisible hand in the cultural evolution of language? A promising future direction is provided by recent attempts to fuse theoretical models of cultural evolution[9] to experimental investigations of social learning in the laboratory[10,11]. Experimental investigations of 'iterated learning' — similar to the game of Chinese whispers, where one participant's output serves as input for the next — can provide empirical data to inspire, and constrain, our theories. Sophisticated new theoretical models enable language-learning 'agents' to have both innate biases (in the form of so-called bayesian priors) and powerful statistical learning systems capable of discovering and using environmental regularities[12]. Such models demonstrate the possibility of a very indirect and sometimes non-intuitive relationship between the regularities emerging at the level of a whole population and the underlying generating forces (Box 1). These forces are individual behaviour and learning (social usage) and innate constraints (in Chomsky's terms, a 'language acquisition device', often called universal grammar).

An important implication of this new synthetic approach to glossogeny is that universals of language are not identical to universal grammar — although obviously related, the two concepts should not be conflated. Another is that cultural evolution can proceed independently of either phylogenetic evolution (the interests of our genes) or our own individual goals and interests. As the papers in this issue[3,4] make clear, cultural evolution can make language easier to learn by filtering out irregular 'noise'[13] — as Lieberman *et al.* wryly point out[3], every "rule is the tombstone of a thousand exceptions". But it can also preserve the irregular cases that make learning a new language difficult: every surviving exception remains a stumbling block for a thousand new language learners.

Nonetheless, as Pagel *et al.* suggest[4], some of the most persistent 'cultural replicators' — memes[14] — evolve as slowly as some genes. By documenting and quantifying such effects, this work opens the door to a diverse range of theoretical and empirical investigations. If there is ever to be a science of memetics[15–17] to rival that of genetics, it should proceed along these lines: combining careful quantitative analysis of well-documented linguistic changes with sophisticated theoretical models capable of taking into account the multilayered complexity of cultural evolution. ■

W. Tecumseh Fitch is in the School of Psychology, University of St Andrews, Fife KY16 9JP, UK.
e-mail: wtsf@st-andrews.ac.uk

1. Chomsky, N. *Knowledge of Language: Its Nature, Origin, and Use* (Praeger, Westport, CT, 1986).
2. Hurford, J. in *Logical Issues in Language Acquisition* (ed. Roca, I. M.) 85–136 (Foris, Dordrecht, 1990).
3. Lieberman, E., Michel, J.-B., Jackson, J., Tang, T. & Nowak, M. A. *Nature* **449,** 713–716 (2007).
4. Pagel, M., Atkinson, Q. D. & Meade, A. *Nature* **449,** 717–720 (2007).
5. Bybee, J. & Hopper, P. (eds) *Frequency and the Emergence of Linguistic Structure* (Benjamins, Amsterdam, 2001).
6. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA, 1949).
7. Heine, B., Claudi, U. & Hünnemeyer, F. *Grammaticalization: A Conceptual Framework* (Univ. Chicago Press, 1991).
8. Keller, R. *On Language Change : The Invisible Hand in Language* (Routledge, New York, 1994).
9. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (Univ. Chicago Press, 1985).
10. Kalish, M. L., Griffiths, T. L. & Lewandowsky, S. *Psychonom. Bull. Rev.* **14,** 288–294 (2007).
11. McElreath, R. *et al. Evol. Hum. Behav.* **26,** 483–508 (2005).
12. Kirby, S., Dowman, M. & Griffiths, T. L. *Proc. Natl Acad. Sci. USA* **104,** 5241–5245 (2007).
13. Deacon, T. W. *The Symbolic Species: The Co-evolution of Language and the Brain* (Norton, New York, 1997).
14. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, 1976).
15. Dennett, D. C. *Darwin's Dangerous Idea* (Simon & Schuster, New York, 1995).
16. Blackmore, S. J. *The Meme Machine* (Oxford Univ. Press, 2000).
17. Mesoudi, A., Whiten, A. & Laland, K. N. *Evolution* **58,** 1–11 (2004).

ORGANIC CHEMISTRY

Zippier synthesis in water

Masayuki Inoue

Complex toxin molecules are the ultimate challenge for organic chemists — even successful syntheses often involve an impractical number of steps. A biologically inspired reaction might simplify things.

Nature creates many complex organic molecules, but some of the most spectacular are the brevetoxins and ciguatoxins produced by marine microorganisms[1]. Brevetoxins are the toxic component of dramatic natural phenomena known as red tides (Fig. 1) — toxic algal blooms that kill huge quantities of fish — whereas ciguatoxins are found in the flesh of fish, and can cause widespread poisoning in humans if they enter the food-chain. Many researchers have invested enormous effort into making these intricate molecules, and have come up with heroic syntheses that stretch the limits of organic chemistry. Reporting in *Science*, Vilotijevic and Jamison[2] now point us towards a potential synthetic short cut. They have found that pure water 'zips up' simple starting materials to make part of the core structures of the molecules, without the need for any other additives.

Brevetoxins and ciguatoxins belong to the family of ladder-like molecules known as polycyclic ethers, which contain fused rings of carbon and oxygen atoms (Fig. 2a, overleaf). A remarkable feature of these molecules is the striking regularity with which the oxygen atoms bridge the carbon framework. In the 1980s it was proposed[3,4] that the key step in the biosynthesis of these compounds is a single 'zip reaction' of a linear precursor molecule into the polycyclic product, catalysed by an unknown enzyme (Fig. 2a). Since then, chemists have tried to emulate such cyclization cascades in the laboratory using a variety of organic solvents, artificial catalysts and specially designed precursor molecules. But these efforts have met with only partial success[5,6].

Such cyclization reactions pose two big problems. First, even for simple cases in which a single ring of atoms is formed, two potential products can be made: a smaller ring of five atoms, or a larger ring of six atoms (Fig. 2b). The cyclization of ten consecutive rings, such as occurs in brevetoxin-A, could therefore theoretically generate 1,024 (2^{10}) products. Designing a reaction that yields only the desired product out of all of these options is challenging, to say the least.

The second, more serious problem stems from the intrinsic preference of cyclization reactions to make one product rather than the other. Organic chemists have known for many years that smaller rings form more readily than their larger counterparts. But assuming that brevetoxin-A is formed from a linear precursor, consecutive 'larger' ring sizes appear in the structure where smaller sizes would be expected. The ring structure of brevetoxin-A could thus theoretically be the least favourable option of all 1,024 possible products. The synthesis of brevetoxin-A in a polycyclization reaction without the aid of an enzyme starts to look unrealistic.

Nevertheless, synthetic chemists have succeeded in reversing the intrinsic selectivity of these cyclization reactions by carefully designing their starting materials. An early example of this involved manipulating the epoxide reacting group in the starting material[7]. Epoxides contain two carbon–oxygen (C–O) bonds; during cyclizations, the product of the reaction — either a five-membered or a six-membered ring — is determined by which of the two bonds breaks. But if an olefin, which contains a carbon–carbon double bond (C=C), is attached to the epoxide, the C–O bond adjacent to the olefin breaks preferentially (Fig. 2b). This is because the olefin partially donates some of its electrons to the adjacent C–O bond, destabilizing the bond so that it ruptures if treated with acid in an organic solvent. Strategic attachment of an olefin to the epoxide can thus direct cyclization reactions to form six-membered rings. This was a crucial synthetic discovery. But it is obviously not how polycyclic ethers are assembled in nature, because no possible brevetoxin precursor uses directing groups such as an olefin.

Vilotijevic and Jamison's method[2] provides a considerable advance over earlier approaches to polycyclic ether formation. They show that a series of three epoxides in a molecule can be reacted in one step to zip together ladder-like structures of six-membered rings, simply by heating them in water (Fig. 2c). The reaction occurs with a good chemical yield at neutral pH, and in the absence of olefin directing groups or other epoxide activating agents.

The reaction mechanism for this unusual transformation[2] has not yet been established, but it is known

B. BACHMAN/SPL

Figure 1 | Crimson tide. Toxic algal blooms wash ashore in Queensland, Australia.

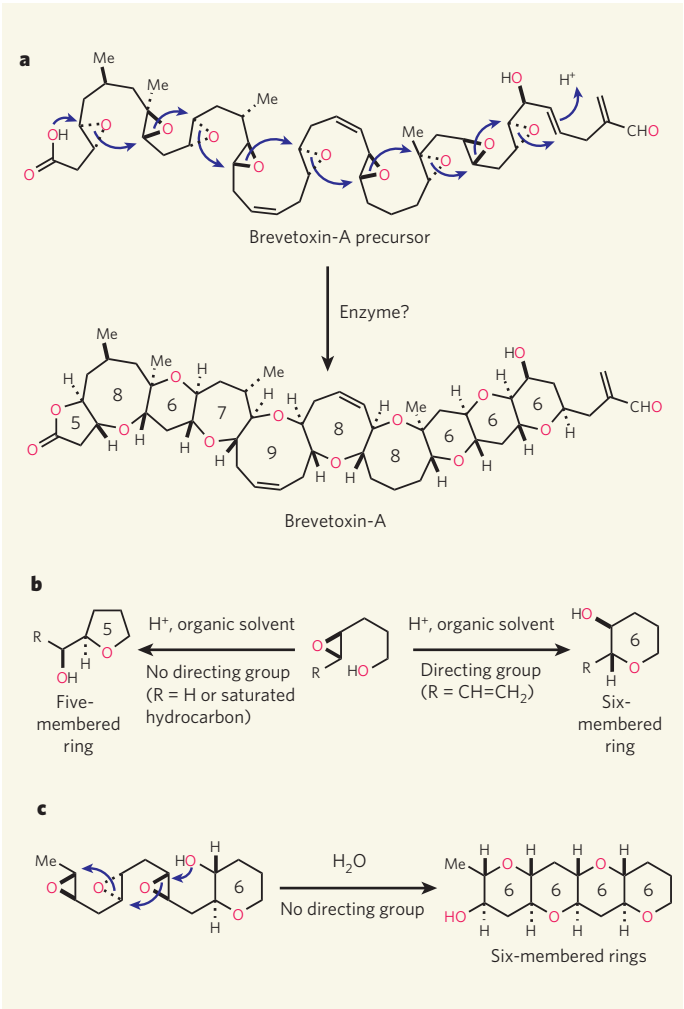


Figure 2 | Synthetic routes to marine toxins. a, Brevetoxin-A is a toxin found in marine algae. The regular arrangement of oxygen atoms in its structure suggests that a key step in its biosynthesis is a 'polycyclization' reaction, probably enzyme-mediated, in which a linear molecule reacts to form many rings of atoms. Curly arrows indicate where bonds form, and the number of atoms per ring is indicated in the product. **b**, In simple model systems of the biosynthesis, unwanted five-membered rings form unless a 'directing' group is attached to the starting material. **c**, Vilotijevic and Jamison[2] now report a polycyclization reaction that mimics the proposed biosynthetic process simply by heating the starting material in water. Six-membered rings form without the need for a directing group.

that water is essential — other solvents are less effective — and that the pH must be neutral to promote six-membered-ring formation. These intriguing observations will undoubtedly stimulate other researchers to re-examine epoxide-opening reactions that have previously been performed in organic solvents.

Vilotijevic and Jamison's results[2] support the hypothesis that molecules such as brevetoxins are made in polycyclization reactions *in vivo*, but they also cast doubt on a common belief: that specific enzymes are necessary for these transformations to occur in nature. So, could pure water alone convert a naturally occurring polyepoxide into brevetoxin-A, in high yield? The answer is probably no. Natural polycyclic ethers, including brevetoxin-A, typically contain rings made of seven, eight and even nine atoms, which are much tougher to construct than six-membered rings. This is because the bonds in the larger rings are more strained, and unfavourable interactions occur between atoms lying across the ring from each other. Even so, the authors' strategy undoubtedly represents the most efficient and sophisticated method for making six-membered-ring ladder structures, which are often found in naturally occurring compounds.

Over the past 20 years, several groups have synthesized marine toxins[8–10], motivated by these compounds' challenging molecular architectures and potent biological activity. Such work is at the cutting edge of organic chemistry, but no genuinely practical method has been devised for making such compounds or their structurally related analogues. The existing syntheses[8,9] — which do not use single-step polycyclization reactions — typically require more than 100 separate reactions. Vilotijevic and Jamison's reaction[2] could allow much shorter routes to be devised. Nevertheless, the development of more practical and general one-step zip reactions remains a key challenge for the future. It might finally allow these compounds to be prepared in sufficient quantities for their biological mechanism of action to be studied. ■

Masayuki Inoue is at the Graduate School of Pharmaceutical Sciences, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
e-mail: inoue@mol.f.u-tokyo.ac.jp

1. Yasumoto, T. & Murata, M. *Chem. Rev.* **93,** 1897–1909 (1993).
2. Vilotijevic, I. & Jamison, T. F. *Science* **317,** 1189–1192 (2007).
3. Lee, M. S., Qin, G., Nakanishi, K. & Zagorski, M. G. *J. Am. Chem. Soc.* **111,** 6234–6241 (1989).
4. Prasad, A. V. K. & Shimizu, Y. *J. Am. Chem. Soc.* **111,** 6476–6477 (1989).
5. Fujiwara, K. & Murai, A. *Bull. Chem. Soc. Jpn* **77,** 2129–2146 (2004).
6. Valentine, J. C., McDonald, F. E., Neiwert, W. A. & Hardcastle, K. I. *J. Am. Chem. Soc.* **127,** 4586–4587 (2005).
7. Nicolaou, K. C., Prasad, C. V. C., Somers, P. K. & Hwang, C. K. *J. Am. Chem. Soc.* **111,** 5330–5334 (1989).
8. Nicolaou, K. C. *et al. Nature* **392,** 264–269 (1998).
9. Hirama, M. *et al. Science* **294,** 1904–1907 (2001).
10. Inoue, M. *Chem. Rev.* **105,** 4379–4405 (2005).

LASER PHYSICS

A phase it's going through

David S. Citrin

When we observe laser light, we typically measure its intensity, and so wave amplitude. The phase, which encodes further details of the laser's internal workings, was obscure — but fresh light is being shed on it.

Watch waves on the ocean. They can be characterized by their height (amplitude) and by when the peaks arrive (phase); without amplitude and phase, there is no wave. Light is also a wave, but one that is composed of intimately linked electric and magnetic fields. Most of the time, we can infer from experiment only the amplitude of light waves: this quantity is given by the intensity averaged over time, which is easily measured.

The phase of the wave is, by contrast, often ignored. But it, too, contains subtle, but no less important, information about the wave and the physical mechanisms that produced it. Knowing exactly what the phase is doing involves following dynamics that may be too fast to track directly, quite unlike ocean waves, which are easy to track with the eye. Typically, our only handle on the phase at the frequencies of optical light has been the interference patterns cast by the light waves. Not any more: on page 698 of this issue[1], Kröll *et al.* directly view both the amplitude and the phase of light in an operating semiconductor laser.

Their laser is a quantum cascade laser[2], which is composed of numerous layers of two or more semiconductors. Under the influence of a d.c. electrical voltage (known as a bias), electrons in atoms of the semiconductor material fall down a staircase of quantum-mechanical levels, giving out a photon at each step (Fig. 1). One of the most exciting aspects of

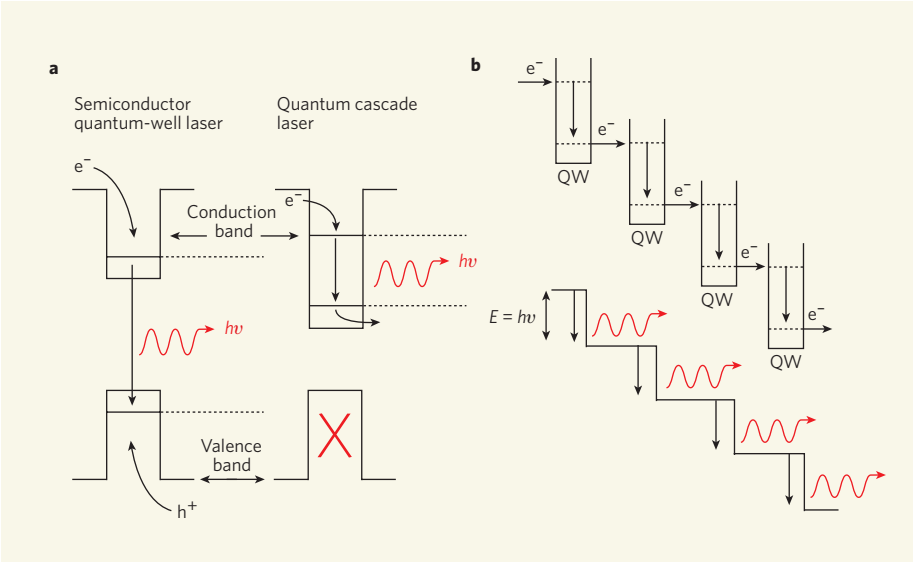


Figure 1 | Light from a quantum cascade. a, In a conventional semiconductor quantum-well laser, an electron undergoes a transition between a quantum-mechanical state in the conduction band and a state in the valence band. This can be viewed as a recombination of an electron (e^-) with a hole (h^+). In the process, a photon is emitted whose frequency is determined by the energy difference between the initial and final electron states. In the quantum cascade laser, the transition is between two quantum-mechanical levels (usually) within the conduction band. **b**, When many quantum-well (QW) structures are stacked up in a quantum cascade laser, the lower level of one stage feeds into the upper level of the next. An electron cascades through the structure, producing as many photons as there are stages.

quantum-cascade approaches is that the light thus emitted has very low frequencies that cannot be produced using conventional lasers. The lowest frequency demonstrated so far is about 1.6 terahertz (THz, 10^{12} Hz). This is close to the clock rates common in high-speed electronics, and so opens up the tantalizing prospect of a further convergence between photonic and electronic technologies. Conventional lasers, by comparison, operate at near-infrared or visible frequencies. This difference translates into oscillation periods of around a picosecond (10^{-12} s) for terahertz quantum cascade lasers, as against around a femtosecond (10^{-15} s) for visible or near-infrared lasers. On terahertz timescales, the task of resolving in time both amplitudes and phases is thus far easier.

The main difference between a normal semiconductor and a laser is that, if the frequency of a light beam shone on the semiconductor is above the fundamental absorption edge (the frequency needed to promote an electron from the valence band to the conduction band), the intensity of the light falls off exponentially as it passes through the medium and its energy is absorbed. A laser turns this picture on its head: light actually extracts energy from the medium. This energy is stored as an inverted population — that is, many electrons reside at a higher-energy level, whereas a lower-energy state is largely empty. When electrons in the upper state undergo quantum-mechanical transitions to the lower state, a photon is emitted whose energy (and so frequency) is given by the difference between the energies of the two levels. Initially, light seeded by this 'spontaneous emission' grows exponentially in intensity as the light propagates in the laser's medium and picks up more and more photons.

But a laser is more than an optical-gain medium: it exploits feedback through the light being partially reflected by the mirrors at either end of the gain medium to build up light intensity to large values. If the net light amplification exceeds light losses through various processes of scattering and absorption, there is net gain, and the structure sustains laser oscillations. The relatively small, though often very intense, amount of coherent light that escapes through an end mirror is the resultant 'stimulated emission' — laser light, as we know it. Inside the laser cavity, however, the light intensity can be much greater.

So far, there has been no word about phase; these simple ideas just balance energy or power. Time-averaged optical power is directly related to the amplitude of the light, not the phase. But phase makes all the difference. The laser's gain medium is in effect a collection of 'dipoles' that are oscillating (think electrons bouncing on springs) at the same frequency as the probe light and themselves spontaneously emitting light. If they are oscillating out of phase with the injected light, they will interfere destructively and attenuate it; this is absorption. But if they are in phase, constructive interference ensues — amplification.

How, then, can we find a window on these inner workings of an operating semiconductor laser? Kröll and colleagues' approach[2] involves injecting a separate pulsed terahertz wave into the gain medium of a quantum cascade laser operating at about 2.9 THz. When this probe pulse leaves the laser again, both its amplitude and phase are analysed. The resulting distortions in this 'temporal shape' provide a unique view of what is going on inside the laser.

The authors observe that, at zero electrical bias, the medium absorbs a range of frequencies contained in the probe pulse. As the bias is increased, but is still below the lasing threshold, gain develops as the quantum-mechanical levels between successive cascades are brought into line. This process allows electrons to occupy the states necessary to build up population inversion, and the distortions suffered by the probe pulse indicate how the light is amplified at what will be the lasing frequency. This distortion shows up directly as a phase flip in the portion of the terahertz pulse that is at the laser frequency.

Up to this point, the quantum cascade laser is operating like an optical amplifier, exhibiting amplified spontaneous emission. But when the bias is increased beyond the laser threshold, stimulated emission builds up substantially. Kröll *et al.* resolve this stimulated emission in time by performing a Fourier analysis on the distorted terahertz pulses, breaking them down into their constituent frequencies. The authors observe the build-up of the pulse as it makes successive trips back and forth through the laser cavity, seeding the laser process, and is further amplified by the medium; and they see 'gain clamping', when the medium cannot supply any more gain.

The techniques developed by Kröll *et al.* can be used to correlate the distortions in the probe pulses with various fast perturbations inflicted on the laser. The work promises a series of studies of laser dynamics and coherent nonlinear optical properties that has hitherto not been feasible. Knowing how a laser recovers its steady state following perturbation will allow fundamental aspects of the underlying physics of the laser to be revealed[3]. Moreover, similar experiments might shed light on ultrafast nonlinear phenomena[4], such as the recent observation of coherent instabilities related to 'Rabi oscillations' that occur in quantum cascade lasers[5] — but not in semiconductor lasers of other types.

The importance of a fuller understanding of the factors that limit gain, cause losses and govern nonlinear optical processes in lasers cannot be overstated. Future generations of terahertz devices with higher power output, lower operating frequencies and ultrafast pulsed operation will find work in a range of applications, including spectroscopy, commerce and security. ■

David S. Citrin is in the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250, USA, and the Unité Mixte Internationale 2958 Georgia Tech-CNRS, Metz, France.
e-mail: david.citrin@ece.gatech.edu

1. Kröll, J. *et al. Nature* **449,** 698–701 (2007).
2. Faist, J. *et al. Science* **264,** 553–556 (1994).
3. Choi, H. *et al.* Abstr. CTuO3 *Proc. Conf. Lasers Electro-Optics (CLEO)*, Baltimore, MD (2007).
4. Gmachl, G. *et al. IEEE J. Quant. Electron.* **39,** 1345–1355 (2003).
5. Wang, C. Y. *et al. Phys. Rev. A* **75,** 31802–31805 (2007).

CANCER

Micromanagement of metastasis

Patricia S. Steeg

Although they were discovered only in the early 1990s, many regulatory functions of microRNAs — naturally occurring short RNA sequences — have already been reported. The latest news is that they mediate cancer spread.

To successfully spread, or metastasize, a tumour cell must complete a complex set of processes, including invasion, survival and arrest in the circulatory system, and colonization of foreign organs[1]. How are these events regulated? In a paper published in this issue (page 682), Ma *et al.*[2] propose that microRNAs (miRNAs), which regulate levels of messenger RNAs (mRNAs), coordinate some of the intricate gene-expression programmes implicated in cancer metastasis.

Since the first description of their regulatory role in the development of the nematode worm *Caenorhabditis elegans* in 1993, more than 400 miRNAs have been discovered. About 1,000 are predicted to exist in humans, regulating the expression levels of roughly 30% of gene transcripts[3]. In human tumours, miRNAs are widely deregulated, and functional studies indicate that they regulate differentiation, proliferation and programmed cell death. The work of Ma *et al.* adds tumour metastasis to the growing list of miRNA-regulated cellular functions.

Few cellular processes have been identified that are more complicated than tumour metastasis. Just imagine a tumour cell's movement. The leading edge of the cell protrudes as a result of cytoskeletal alteration and adherence to the extracellular matrix, whereas the trailing edge detaches from the matrix. We do not know how many separate gene-expression programmes are initiated in simple motility, let alone the many additional requirements of metastasis. But miRNAs are prime candidates for coordinating these complex gene-expression events that lead to metastasis.

Earlier work[4] that compared miRNA-expression patterns in healthy breast tissue and primary breast carcinomas found that several miRNAs are expressed differently in cancerous cells. Taking advantage of this observation, Ma and colleagues[2] asked which of these miRNAs are differentially expressed in metastatic breast-cancer cells but not in healthy or in non-metastatic tumorigenic cells. They identified miR-10b, the expression of which has previously[4] been shown to correlate with vascular invasion. The authors also found that miR-10b specifically augments tumour-cell invasion without affecting the cells' viability or proliferation.

To explore the relevance of these findings to cancer spread *in vivo*, the authors investigated the effect of miR-10b on metastasis in mice. They used two different breast-cancer cell lines, manipulating each to express increased levels of miR-10b. Injection of one line into mice caused the animals to develop pulmonary micrometastases — clusters of tumour cells. With the second cell line, 80% of mice showed visible signs of lung metastases, and in 30% cancerous cells metastasized to the peritoneum — the membrane covering the wall of abdominal cavity. These findings clearly show that miR-10b regulates breast-tumour metastasis, but that its impact depends on other factors — in this case, cellular context.

A fascinating aspect of this work is the elucidation of the mechanism through which miR-10b potentially regulates metastasis. The authors found that Twist, a gene-transcription factor that has been implicated in the regulation of metastasis[5], induces miR-10b transcription in human mammary epithelial cells. It does this by binding to a specific DNA sequence known as E box, which is located upstream of the sequence leading to miR-10b. Thus, miR-10b seems to be under the direct control of at least one transcription factor.

To repress translation or facilitate degradation of mRNAs, miRNAs bind to their 3′ untranslated regions[3]. The authors[2] used a computational model to search for potential miR-10b binding sites and identified about 100 potential targets. Of these, they investigated the relevance of the HOXD10 protein, which had been implicated in the suppression of tumour-cell motility. They found that miR-10b inhibits the translation of HOXD10 mRNA, thereby affecting the expression of the downstream targets of this transcription factor. One effector of HOXD10 is the RHOC protein, which functions in molecular signalling cascades mediating metastasis[6,7]. Indeed, the authors found that miR-10b increased RHOC expression.

Put together, the findings of Ma *et al.* join the dots in the miR-10b-mediated signalling pathway that regulates cancer metastasis (Fig. 1). The transcription factor Twist stimulates miR-10b expression, which reduces the levels of another transcription factor, HOXD10. This results in increased levels of RHOC, which stimulates cancer-cell motility. However, not all metastasis-promoting activities of Twist, such as the transition of epithelial cells to mesenchymal cells — a re-differentiation programme involved in cancer-cell invasion — could be ascribed to miR-10b. So these results identify only one Twist-mediated pathway of importance to motility — and potentially metastasis — which is regulated by miR-10b.

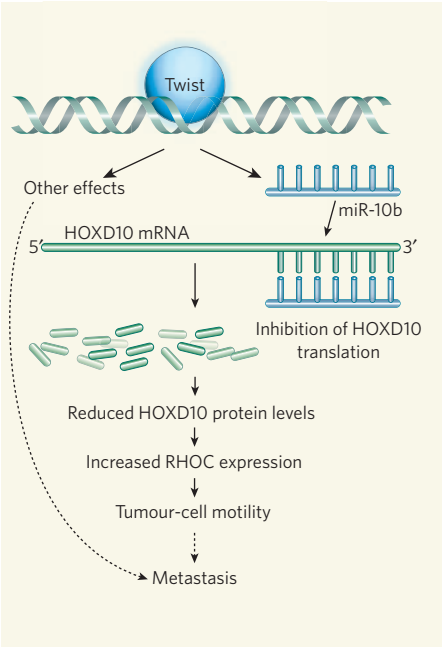


Figure 1 | The microRNA-mediated pathway implicated in metastasis. Ma *et al.*[2] investigate the complex role of the transcription factor Twist in tumour-cell motility. They find that Twist induces the transcription of microRNA-10b (miR-10b), which inhibits translation of, among other sequences, the mRNA for the transcription factor HOXD10 by binding to its 3′ untranslated region. This, in turn, increases both the expression of RHOC and cell motility. Not all of Twist's known downstream effects leading to metastasis are regulated by the miR-10b-mediated pathway.

But how relevant is this pathway to the metastasis of human cancers? The authors found that, in all five metastasis-free cancer patients they examined, miR-10b expression in tumours was lower than in associated normal tissue. However, the levels of this miRNA were elevated in the tumours of 9 out of 18 metastasis-positive patients. Although not definitive, this observation will undoubtedly stimulate further clinical studies.

Other miRNAs might also contribute to metastasis. For example, the miR-17–92 cluster of miRNAs regulates the expression of connective-tissue growth factor and thrombospondin-1 (ref. 8), which are important for angiogenesis — the formation of new blood vessels. Candidate miRNA-regulated mRNAs whose expression is thought to mediate metastasis have also been suggested[9]; these include mRNAs for proteins involved in different aspects of metastasis, such as adherence, invasion, motility, angiogenesis and protein degradation. A relevant question is, therefore, whether individual miRNAs can contribute to several events associated with metastasis and, if so, how they are regulated in a cellular context. If different miRNA pathways consistently control various aspects of metastasis, therapeutic strategies for targeting miRNAs that have been implicated in

other pathological processes[10] may be applicable to preventing cancer spread. ■

Patricia S. Steeg is at the National Cancer Institute, Building 37, Room 1122, 37 Convent Drive, Bethesda, Maryland 20892, USA.
e-mail: steegp@mail.nih.gov

1. Steeg, P. S. *Nature Med.* **12,** 895–904 (2006).
2. Ma, L., Teruya-Feldstein, J. & Weinberg, R. A. *Nature* **449,** 682–688 (2007).
3. Kent, O. A. & Mendell, J. T. *Oncogene* **25,** 6188–6196 (2006).
4. Iorio, M. V. *et al. Cancer Res.* **65,** 7065–7070 (2005).
5. Yang, J. *et al. Cell* **117,** 927–939 (2004).
6. Clark, E. A., Golub, T. R., Lander, E. S. & Hynes, R. O. *Nature* **406,** 532–535 (2000).
7. Hakem, A. *et al. Genes Dev.* **19,** 1974–1979 (2005).
8. Dews, M. *et al. Nature Genet.* **38,** 1060–1065 (2006).
9. Dalmay, T. & Edwards, D. R. *Oncogene* **25,** 6170–6175 (2006).
10. Hammond, S. M. *Trends Mol. Med.* **12,** 99–101 (2006).

EVOLUTIONARY GENETICS

Making the most of redundancy

Edward J. Louis

Single genes, chromosomal regions and even entire genomes can undergo duplication. What good can come of these extra copies? Evolution seems to use several tricks to take advantage of the situation.

Gene mutations often result in abnormal levels or function of their protein products. Consequently, the divergence rate of DNA sequences that encode genes is generally slower than that for non-coding sequences. So how does new genetic material arise? One valuable source is sequences formed through gene and genome duplication events.

The commonest consequence of genomic duplication seems to be the loss of all or part of duplicated sequences through deletion[1] or degeneration[2], causing non-functionality. This can be a powerful evolutionary force — for example, gene loss in response to whole-genome duplication led to rapid speciation in yeast[3]. Alternatively, although it is much less-well understood, one or both of the duplicated sequences might acquire new functions, as they are under less selective pressure, and can afford to undergo mutations that would lead to new characteristics and functions. On page 677 of this issue, Hittinger and Carroll[4] analyse one such pair of duplicated genes with divergent functions.

Other than by deletion or degeneration following genomic duplication (Fig. 1a), new functional sequences arise mainly through differential degeneration of the two copies, so that the newly evolved functions mutually complement the original, pre-duplication function[5]. This is particularly common in multifunctional genes, and may occur in one of three ways. The two copies can evolve different functions, which may be overlapping or non-overlapping (Fig. 1b), allowing complementation and maintenance of the original function, as well as expansion of functionality. Alternatively, in bifunctional or multifunctional genes, in which different functions are encoded by different sequence modules within the same gene, duplicated copies can reciprocally lose or change function in different modules; this could also result in the coverage of original function and its possible expansion (Fig. 1c). Finally, regulatory differentiation may occur, allowing differential spatial or temporal expression of one or both copies (Fig. 1d). New functions probably arise through a combination of these processes.

One of the most exciting natural experiments in genome evolution is the duplication of whole genomes in *Saccharomyces* yeast species[1]. Comparative genomic analysis[6] of several sequences from these species has revealed some general principles underlying the evolution of duplicated genes. For example, new functions rarely seem to arise from changes in biochemical properties but are frequently associated with regulatory changes.

Hittinger and Carroll[4] investigate the evolution of one pair of duplicated genes in *Saccharomyces cerevisiae*, and conclude that they diverged following a combination of differentiation and degeneracy in different modules, including their regulatory regions. They study the *GAL1* gene, which encodes the galactokinase enzyme, and the *GAL3* gene, which encodes the Gal3 protein. A co-inducer of galactokinase, Gal3 sequesters a repressor of the gene-transcription factor that activates the galactose-use pathway[7]. *GAL1* and *GAL3* evolved from a single bifunctional gene; of these, *GAL3* has lost its enzymatic activity, whereas *GAL1* has changed its regulatory requirements[1]. The outcome is a much more tightly controlled and highly inducible *GAL1*.

Hittinger and Carroll set out to investigate the evolution of *GAL1* and *GAL3* by testing fitness differences in a yeast species, *Kluyveromyces lactis*, which, unlike *S. cerevisiae*, has not undergone whole-genome duplication. In *K. lactis*, the functions of galactokinase and Gal3 are encoded by a single gene, which has one regulatory region. The authors find that increased expression of the gene module providing galactokinase activity in *K. lactis* enhances fitness, whereas overexpression of the module equivalent to *GAL3* reduces fitness. This leads to a conflict. In *S. cerevisiae*, however, there is no such conflict. So the outcome of this adaptive evolution is optimal regulation of the gene encoding the first enzyme in the galactose-use pathway — galactokinase — without maladaptive regulation of the co-inducer of the pathway, Gal3.

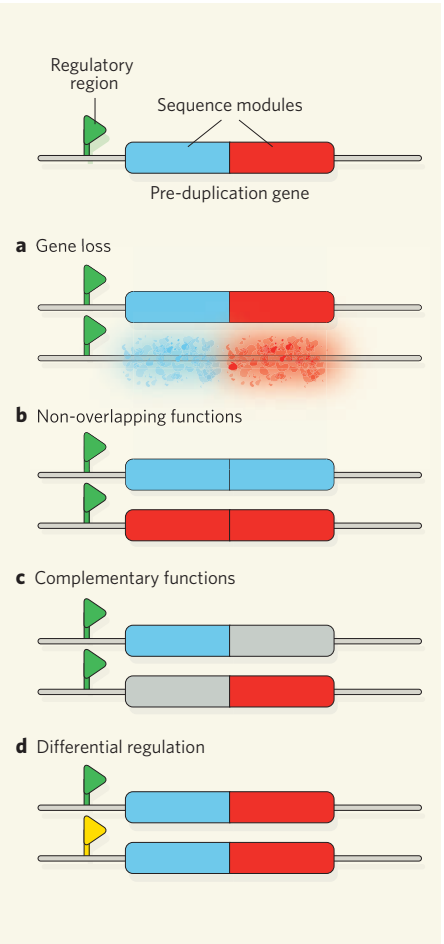


Figure 1 | Possible outcomes of gene duplication. Evolution has devised many ways to cope with, and make good use of, gene duplication. **a**, A common outcome is loss of one of the two copies of the duplicated gene by deletion or degeneration. **b**, Alternatively, the two copies of a gene can diverge in sequence, either maintaining the ability to complement or substitute for each other, or diverging completely to have non-overlapping functions. **c**, A third possibility is that different functional modules of the two copies are lost or diverge, so that together the products of the two copies perform the original function but each is responsible for a single aspect of the function. **d**, Finally, the regulatory regions of the two copies can diverge, allowing spatial and temporal expansion of functionality, as well as enhanced expression levels. The genes Hittinger and Carroll[4] analysed have undergone a combination of the last two of these processes.

Another recent study[8] investigated the duplication of the genes encoding the histone deacetylase enzymes Hst1p and Sir2p in *S. cerevisiae*. The results indicated different evolutionary principles from those observed for *GAL1* and *GAL3*. In this case, the authors

find that although Hst1p and Sir2p have different functions, they use the same biochemical process for their activity. The functional differentiation of these two enzymes is due to their distinct protein-interaction domains, which allow the selection of different protein partners and complex associations. Consequently, Hst1p is involved in the repression of genes expressed in meiotic cell division and Sir2p functions in a process known as silent mating-type regulation. After gene duplication, Sir2p seems to have acquired the ability to interact with new protein partners, leading to the expansion of its function, which now allows it to inactivate chromatin (complexes of DNA and histone proteins).

Although the functions of Hst1p and Sir2p do not overlap in normal cells, if the gene encoding Hst1p is deleted, Sir2p, which has some affinity for the protein partner of Hst1p, maintains genetic robustness, thereby partly substituting for Hst1p function. The pre-duplication species in the case of genes encoding Hst1p and Sir2p is also *K. lactis*, which the authors[8] studied next. They found that the single pre-duplication gene in this yeast species can both complement the enzymatic activity of Hst1p and interact with the protein partner of Sir2p when expressed in *S. cerevisiae*. Moreover, analysis of chimaeras of the genes from the pre- and post-duplication species indicated that it is the evolution of distinct protein-interaction domains that has allowed this differentiation of functionality.

Unravelling the processes that underlie the evolution of new functions after gene duplications will be an important area of research for some time[6]. The new era of comparative genomics, together with the sequencing of an ever-increasing number of genomes, will provide a great deal of data for testing various hypotheses about how these processes occur.

Furthermore, in yeasts we have the ideal model organisms, because the genomes of many yeast species have been sequenced. In addition, there are many yeast species that have undergone whole-genome duplication, and others that have instead suffered smaller-scale segmental duplications. These analyses are relevant because other taxa and, in particular, multicellular organisms, have also undergone many large- and small-scale duplication events, which must have offered many opportunities for regulatory and functional differentiation.

With numerous natural experiments to analyse, we have many opportunities that would not have been possible before the advent of large-scale genome sequencing and systems-biology approaches. We now have the chance, for example, to work out the underlying processes that can lead to changes in gene function, and perhaps to discover processes and principles that will help us to understand the form and function of genes. This view is, of course, controversial. Some are adamant that we have all the theoretical tools necessary to explain life's complexity[9], whereas others maintain that we might be missing central concepts and principles[10]. Regardless of whether we know the unknowns or whether there are unknown unknowns, genomic redundancy offers exciting research opportunities. ■

Edward J. Louis is at the Institute of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK.
e-mail: ed.louis@nottingham.ac.uk

1. Wolfe, K. H. & Shields, D. C. *Nature* **387,** 708–713 (1997).
2. Koszul, R., Caburet, S., Dujon, B. & Fischer, G. *EMBO J.* **23,** 234–243 (2004).
3. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. *Nature* **440,** 341–345 (2006).
4. Hittinger, C. T. & Carroll, S. B. *Nature* **449,** 677–681 (2007).
5. Force, A. *et al. Genetics* **151,** 1531–1545 (1999).
6. Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. *Nature* **449,** 54–61 (2007).
7. Johnston, M. *Microbiol. Rev.* **51,** 458–476 (1987).
8. Hickman, M. A. & Rusche, L. N. *PLoS Genet.* **3,** e126 (2007).
9. Lynch, M. *The Origins of Genome Architecture* (Sinauer, Sunderland, MA, 2007).
10. Carroll, S. B. *Nature* **409,** 1102–1109 (2001).

SOLID-STATE PHYSICS

Response with a twist

Karin M. Rabe

The behaviour of ferromagnetic and ferroelectric materials in a magnetic or electric field makes them easy to spot. But for their more recently discovered counterpart, ferrotoroidic materials, things become complex.

Ordering, with its effects on the symmetry and properties of crystals, is central to many phenomena in solid-state physics. The most familiar example is ferromagnetism: the alignment, even in the absence of an applied magnetic field, of local magnetic moments in a material that results in a uniform, permanent magnetization. Indeed, the use of the prefix 'ferro' to refer to this type of uniform ordering derives from the historical observation of permanent magnetization in metallic iron. Analogously, in ferroelectrics, atomic displacements produce a uniform electric polarization even in the absence of an applied electric field.

On page 702 of this issue, Van Aken *et al.*[1] take a step towards adding a new member to this family of orderings: 'ferrotoroidicity'. Just as the ability to control the magnetic and electrical properties of ferromagnets and ferroelectrics has put these materials at the heart of many modern technologies, most notably in information storage, so we might expect useful applications for ferrotoroidic materials once this form of ordering has been fully characterized and a range of materials exhibiting it has been identified.

Ferrotoroidicity is particularly intriguing in that it combines symmetry-breaking features of both ferromagnetism and ferroelectricity. Ferromagnetism depends purely on the orientation of magnetic moments, which is an expression of the direction of electron spins and their orbital angular momentum. When the direction of time is reversed, the electrons spin the other way and the magnetization thus changes sign; but under inversion symmetry, when the spatial coordinates are flipped, it remains unchanged. Ferroelectricity, on the other hand, is produced by a spatial ordering of charges that is static in time. It is therefore unaffected by time reversal, but changes sign under inversion (Fig. 1).

The presence of ferrotoroidic order is quantified by a vector, called the toroidization, that expresses the formation and uniform alignment of local vortices of magnetic moments. These vortices are defined by both the orientation of the magnetic moments and their spatial arrangement on the crystal lattice[2]. Thus, the sense of these vortices reverses under both time reversal and spatial inversion, with a corresponding reversal of the toroidization (Fig. 1c). In analogy to ferromagnetic and

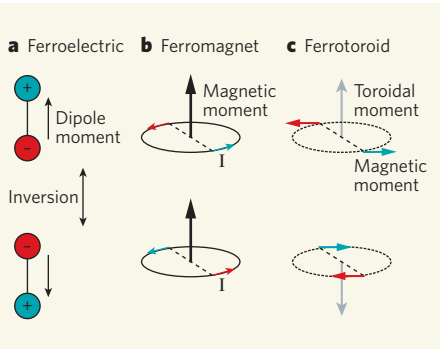


Figure 1 | Topsy-turvy world. a, Inversion — the reversal of spatial coordinate axes — interchanges the positive and negative charges in a dipole, and so reverses the sign of the dipole moment: ferroelectric ordering of charges thus breaks inversion symmetry. **b**, Ferromagnetic ordering, on the other hand, depends on the alignment of local magnetic moments, which are produced by electrons' spins and orbital angular momenta. The magnetic moment of a classical current loop demonstrates the symmetry: inversion maps the current (I) vector at each point of the loop to its antipode and also reverses its direction, leaving the magnetic-moment vector unchanged. **c**, In the vortex of magnetic moments found in a ferrotoroid, inversion maps each of the two moments to its antipode, keeping its direction unchanged. The sense of the vortex, and thus the sign of the toroidal moment, is reversed.

ferroelectric materials, two adjacent regions of the crystal can choose opposite directions for the toroidization, forming ferrotoroidic domains separated by a well-defined wall.

It is this presence of adjacent regions with opposite toroidization that makes the detection of ferrotoroidic order possible: measuring the difference in toroidization between two states is, in principle, not as hard as measuring absolute values. Indeed, as in the case of electric polarization for an infinite periodic crystal[3,4], a rigorous definition of the absolute magnitude of the toroidization of an infinite crystal proves to be rather subtle[2].

Van Aken and colleagues' contribution[1] is to establish, for the first time, the presence of ferrotoroidic order by detecting the difference in the phase of light bounced off adjacent domains of opposite toroidizations. Their chosen proving ground is the oxide compound $LiCoPO_4$, which, on the basis of previous studies of its crystal and magnetic structure[5], is expected to exhibit ferrotoroidic order. It contains planes of CoO_6 octahedra, buckled by the sharing of corners with PO_4 tetrahedra above and below the plane (Fig. 2). The local magnetic moments of cobalt cations are ordered antiferromagnetically, changing sign from row to row in the plane, and are oriented along a direction tilted by 4.6° from the axis of the row. The non-zero toroidization relies on both the buckling of the planes and the tilt of the magnetic moments.

For the optical imaging of ferrotoroidic domains, regions with opposite toroidizations must scatter light in a way that makes them distinguishable both from each other and from non-ferrotoroidic crystals. In $LiCoPO_4$, the low symmetry of the ferrotoroidically ordered state means that the electric polarization responds nonlinearly to an applied electromagnetic field, such that some of the light emerging from the crystal has a frequency double that of the incident light. The polarization of the frequency-doubled outgoing light also depends in a characteristic way on the type of ordering, as well as on the polarization of the incident light. Moreover, the relative phase can change depending on the direction of the ordering vector, producing interference effects. In this way, it is possible to image both antiferromagnetic domains[6] and the independent ferrotoroidic domains in $LiCoPO_4$.

The recent surge of interest in ferrotoroidic materials has been driven by the recognition that systems in which magnetic and electric orderings are coupled have physical properties beyond the simplest ones, in which an applied field causes a change in a material property of the corresponding class. Examples of these well-studied simple processes include an electric field acting on charges to produce an electric polarization (dielectric susceptibility) and a magnetic field aligning spins and currents to produce a uniform magnetization (magnetic susceptibility). But 'magnetoelectric' materials, in which an applied electric field produces a uniform magnetization and an applied magnetic field produces an electric polarization, are a different and more exciting proposition altogether[7].

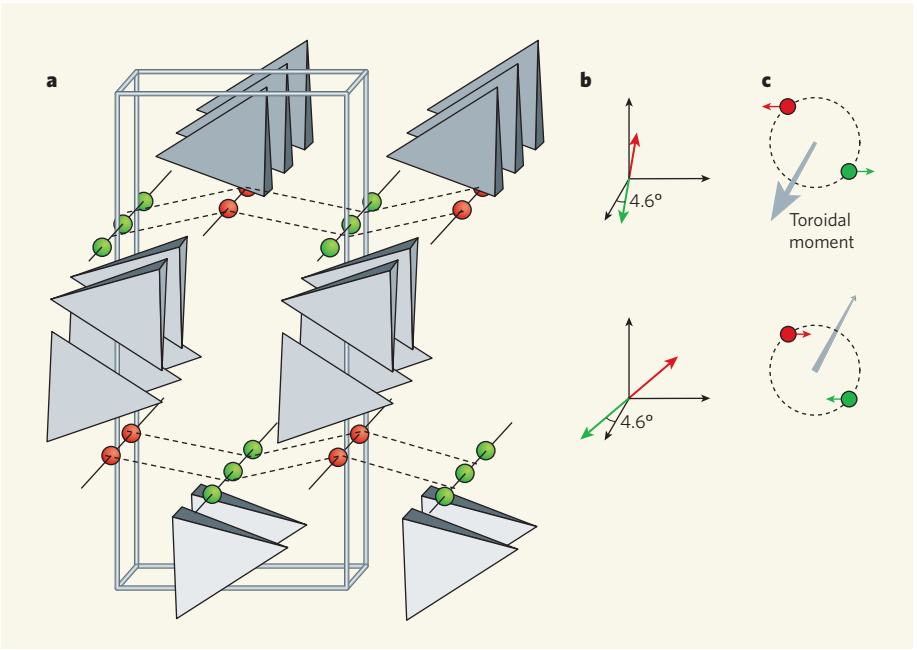


Figure 2 | Ferrotoroidicity in $LiCoPO_4$. **a**, In the $LiCoPO_4$ crystal structure, cobalt ions (spheres) form planar layers, with the magnetic moments ordered antiferromagnetically — that is, pointing in opposite directions in alternate rows. These layers are buckled by the presence of phosphorus-centred oxygen tetrahedra (grey) above and below them (the lithium ions between the tetrahedra are omitted for clarity). **b**, In upward-buckled rows (red), the magnetic-moment vectors of the cobalt ions are tilted at an angle of 4.6° to the row, and the magnetic moments of ions in downward-buckled rows (green) are tilted by the same amount in the opposite direction. **c**, The components of the magnetic moments perpendicular to the rows give rise to the non-zero toroidization out of or into the page.

The essential criterion for a non-zero magnetoelectric effect is that the crystal symmetry should be low. In addition, low symmetry produced by a symmetry-breaking distortion of a higher-symmetry structure gives rise to couplings that are, in principle, tunable. Non-zero toroidization, with its coupling of atomic arrangement and magnetic moments in the vortex, provides enough symmetry breaking to lead naturally to the substantial magnetoelectric coupling that had previously been observed in $LiCoPO_4$ (ref. 8). This contrasts with the generally small magnetoelectric coupling in ferromagnetic ferroelectric (or multiferroic) systems[9], in which these two independent orders combine to break the required symmetry without directly coupling the magnetization and electric polarization at the atomic scale.

Analogies to ferromagnetism and ferroelectricity, and associated couplings such as piezoelectricity (polarization created by mechanical stress), suggest that much may be possible with ferrotoroidically ordered systems. Ferrotoroidic order is a useful conceptual tool for organizing the relevant information about the low-symmetry states of these complex oxide materials. This may allow the tuning of magnetoelectric effects. If it proves possible to produce fields that can switch the sense of ferrotoroidic domains, this might be exploited to flip the sign of particular magnetoelectric tensor coefficients — or even, as with the ferroelectric and ferromagnetic materials that have become standard elements of computer memory devices, to store information in binary states.

A broader perspective will be achieved by comparing ferrotoroidic magnetoelectrics with magnetically induced ferroelectrics[10] and with non-toroidic magnetoelectrics. This could lead to the identification of additional ferrotoroidic materials and a more complete characterization of their behaviour. Crystallographers know well that complex oxides offer a rich variety of structure types, and many possibilities for phase transitions to structural and magnetic states of lower symmetry. As we are now beginning to understand, that can also yield unexpected physical properties. ■

Karin M. Rabe is in the Department of Physics and Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuysen Road, Piscataway, New Jersey 08854-8019, USA.
e-mail: rabe@physics.rutgers.edu

1. Van Aken, B. B., Rivera, J.-P., Schmid, H. & Fiebig, M. *Nature* **449,** 702–705 (2007).
2. Ederer, C. & Spaldin, N. A. preprint available at www.arxiv.org/abs/0706.1974 (2007).
3. King-Smith, R. D. & Vanderbilt, D. *Phys. Rev. B* **47,** 1651–1654 (1993).
4. Resta, R. *Rev. Mod. Phys.* **66,** 899–915 (1994).
5. Vaknin, D., Zarestky, J. L., Miller, L. L., Rivera, J.-P. & Schmid, H. *Phys. Rev. B* **65,** 224414 (2000).
6. Fiebig, M. *et al. Nature* **419,** 818–820 (2002).
7. Fiebig, M. *J. Phys. D* **38,** R123–R152 (2005).
8. Kornev, I. *et al. Phys. Rev. B* **62,** 12247–12253 (2000).
9. Eerenstein, W., Mathur, N. D. & Scott, J. F. *Nature* **442,** 759–765 (2006).
10. Cheong, S.-W. & Mostovoy, M. *Nature Mater.* **6,** 13–20 (2007).

ARTICLES

Gene duplication and the adaptive evolution of a classic genetic switch

Chris Todd Hittinger[1]† & Sean B. Carroll[1]

How gene duplication and divergence contribute to genetic novelty and adaptation has been of intense interest, but experimental evidence has been limited. The genetic switch controlling the yeast galactose use pathway includes two paralogous genes in *Saccharomyces cerevisiae* that encode a co-inducer (*GAL3*) and a galactokinase (*GAL1*). These paralogues arose from a single bifunctional ancestral gene as is still present in *Kluyveromyces lactis*. To determine which evolutionary processes shaped the evolution of the two paralogues, here we assess the effects of precise replacement of coding and non-coding sequences on organismal fitness. We suggest that duplication of the ancestral bifunctional gene allowed for the resolution of an adaptive conflict between the transcriptional regulation of the two gene functions. After duplication, previously disfavoured binding site configurations evolved that divided the regulation of the ancestral gene into two specialized genes, one of which ultimately became one of the most tightly regulated genes in the genome.

Gene duplication has long been recognized as a major source of new genes and functions. Until recently, it was generally assumed that duplicate genes were free to evolve new functions ('neofunctionalization') because the original function was maintained by the other copy[1–5]. However, several recent case studies and comparisons of genome content have suggested that most new genes do not have novel functions[6–10]. Instead, paralogous gene pairs are often 'subfunctionalized' with two or more functions being partially or completely subdivided between the two genes after gene duplication. Various evolutionary models have been proposed to account for the fates of paralogues during subfunctionalization.

The duplication–degeneration–complementation (DDC) model explains the preservation of duplicate genes by a neutral mechanism in which each paralogue accumulates loss-of-function mutations (degeneration) that are complemented by the other copy[11]. Examples of DDC include the complementary losses of spatial domains of expression following gene duplication in multicellular organisms[6,11,12]. DDC may explain the initial retention of most duplicate genes, but DDC alone is not adaptive. Gene duplication is only a necessary contributor to adaptation if it allows for otherwise prohibited adaptive mutations.

One set of circumstances under which some adaptive mutations can be prohibited is in the case of multifunctional genes. The performance of multiple functions by a single gene or 'gene sharing' is prevalent in nature[10] but presents the possibility that mutations that optimize one function may compromise the other. Such an 'adaptive conflict' may be resolved by gene duplication followed by subfunctionalization and positive selection, thereby creating two optimized specialist genes in a manner forbidden during gene sharing[13–15].

Adaptive conflict resolution has been suggested to have been important to the specialization of several duplicated genes[13–15], but demonstration of specific molecular changes involved in adaptive conflict has proven elusive. Here we dissect the evolution of a pair of duplicate genes from the galactose (GAL)-use pathway from their bifunctional ancestor and show that they are nearly completely subfunctionalized in *Saccharomyces cerevisiae*. With one function being performed by its paralogue, *GAL3*, *GAL1* was free to evolve previously disfavoured adaptive changes in its upstream regulatory element that provided a more dynamic and tightly regulated transcriptional response. We propose that an adaptive conflict in the regulation of these two functions was resolved along the lineage leading to *S. cerevisiae* by gene duplication and the subfunctionalization of *GAL1* and *GAL3*.

Gene duplication and the GAL genetic switch. The co-inducer (Gal3) of the classic genetic switch[16–18] of the GAL pathway of *S. cerevisiae* is encoded by a member of a duplicate gene pair created by a whole-genome duplication about 100 million years ago[19–23]. Its paralogue encodes the first enzyme of the GAL pathway, the galactokinase Gal1, the upstream regulatory elements of which are the DNA component of this switch. *GAL1* is one of the most tightly regulated genes in the genome, being repressed in the absence of galactose and induced about 1,000-fold in its presence[24,25]. In contrast, *GAL3* expression is induced a modest three- to fivefold from its basal expression[26]. Gal3 co-induces the pathway by sequestering the co-repressor Gal80 in the cytoplasm in a galactose- and ATP-dependent fashion[27–30]. The transcription factor Gal4 is then free to activate the GAL pathway, including *GAL1*. The switch works similarly in *Kluyveromyces lactis*, a yeast species that did not undergo the whole-genome duplication, except that a single bifunctional *K. lactis* protein encoded by a *GAL1* homologue performs both the co-induction and galactokinase functions[17,27,31,32].

Protein subfunctionalization. Gal1, Gal3 and *K. lactis* Gal1 (which we define hereafter as KlacGal1) are similar in sequence[26,33] and structure[34], and each protein is sufficient to act as the sole co-inducer in plasmid complementation and *in vitro* assays[27,29,35]. Addition of a Ser-Ala (SA) dipeptide to the degenerated active site of Gal3 has been shown to be necessary and sufficient to confer galactokinase activity[35]. These two residues are leading candidates to account for the functional divergence of Gal1 and Gal3[15,17,32,35].

Complementation and *in vitro* assays may demonstrate the sufficiency of a protein to perform a function, but natural selection can operate on very small fitness differences. Optimization of a single subfunction is expected to be quantitative, so more sensitive tests are required to detect functional differences among proteins and

[1]Howard Hughes Medical Institute, Laboratory of Genetics, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA. †Present address: Center for Genome Sciences, School of Medicine, Washington University in St Louis, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA.

possible instances of adaptive conflict. Therefore, we have employed a competitive growth assay capable of detecting fitness differences among strains in the order of 10^{-3} offspring per generation, well beyond the detection limit of previous assays. In this assay, genetically manipulated green fluorescent protein (GFP)-tagged strains are competed against an otherwise identical blue fluorescent protein (BFP)-tagged strain in liquid culture, cells are counted by flow cytometry, and the effect on fitness is calculated. Targeted gene replacement allowed us to make any desired genetic change and provided the control over genetic background necessary to detect small fitness differences. To address the contribution of coding changes to the divergence of these paralogues, we used this assay to measure the fitness effects of targeted replacements of the coding regions of *GAL3* and *GAL1*.

We first examined the ability of various coding sequences to perform the galactokinase function of Gal1 *in vivo*. There was no effect of any coding replacement in non-inducing conditions, but we observed strong fitness defects in galactose when other proteins were required to perform the function of Gal1 (Table 1). This was expected for Gal3 (known to have no galactokinase activity[26,35]) and for proteins designed to mimic the SA dipeptide deletion in the would-be active site of Gal3 (Gal1-ΔSA and KlacGal1-ΔSA). In contrast to complementation assays[27,35], our quantitative assay detected strong fitness defects for even the galactokinases Gal3-∇SA (a version of Gal3 with the SA dipeptide inserted) and KlacGal1. Gal3-∇SA has a much slower enzymatic rate than Gal1[35], and this seems to be a substantial handicap under competitive conditions. The KlacGal1 defect was largely due to toxicity associated with expressing this heterologous protein at high levels in *S. cerevisiae* and not poor enzyme activity itself (Supplementary Table 1).

We next tested the ability of these coding sequences to perform the co-induction function of Gal3. Gal3-∇SA performed poorly as a co-inducer (Fig. 1), a defect not detectable in previous complementation assays[35]. Most significantly, this result raises the possibility that the SA residues could be a source of adaptive conflict between the galactokinase and co-induction functions. For example, tight binding of galactose and ATP by Gal3 in a complex with Gal80 might be inconsistent with efficient galactose phosphorylation. If this were the case, then deleting these two residues from Gal1 and KlacGal1 should improve their co-induction functions and make them function more like Gal3. Instead, we found that deleting these residues further impairs their co-induction function relative to wild-type Gal1 and KlacGal1 (Fig. 1).

Because the removal of the SA residues from Gal1 and KlacGal1 did not improve their co-induction function, the SA residues were not a source of adaptive conflict; nor are the galactokinase and co-induction protein functions inherently antagonistic. Instead, the presence of the SA residues has opposite effects on the co-induction capabilities of Gal1 and Gal3. The differential effect of the SA residues must be due to coding differences at other sites within the proteins. These opposite effects mean that the SA residues are a case of intragenic 'sign epistasis',[36] which is said to be present when the direction of the fitness effect of a mutation depends on residue(s) present at other site(s) within the protein. The SA residues of Gal1/3 are the first case of sign epistasis reported between paralogues, but sign epistasis may be a general feature of duplicated proteins that have diverged in function and accumulated interacting changes.

Despite the divergence between Gal3 and Gal1 (74% identity) or Gal3 and KlacGal1 (58% identity) in addition to the SA dipeptide, Gal1 and KlacGal1 were fairly effective at performing the co-induction function of Gal3 at higher galactose concentrations (≈2–3% fitness defect; Fig. 1). Even under more challenging conditions of low galactose concentrations, strains expressing Gal1 and KlacGal1 from the endogenous *GAL3* upstream regulatory elements (hereafter, the *GAL3* promoter or P_{GAL3}) still performed relatively well. The performance of these strains contrasts sharply with the *gal3-Δ* strain where Gal1 is driven only by its native *GAL1* promoter (Fig. 1). Taken together, these experiments suggest that, aside from the complete degeneration of the galactokinase activity of Gal3, most of the functional divergence between these paralogues has been regulatory. This inference led us to focus on the upstream regulatory elements as a potential source of adaptive conflict.

Most adaptive divergence is regulatory. To assess the divergence in the upstream regulatory elements, we swapped the promoters of the two paralogues and assessed their capabilities to drive proper expression of *GAL1* and *GAL3*. Competition assays revealed that P_{GAL1} and P_{GAL3} were both inferior at performing the function of the other promoter (Fig. 2; a competition assay in which the tested promoter drives the expression of the paralogous coding sequence in a null mutant background). The inability of P_{GAL3} to drive sufficient levels of galactokinase is due to its comparatively low messenger RNA expression level when fully induced (Table 2)[26]. In contrast, whereas P_{GAL1} is strongly induced in the presence of galactose, it is more

Table 1 | Effects of the targeted replacement of the *GAL1* coding region

Coding region replacing GAL1	Growth	Fitness defect	s.d.
GAL1$^+$	+	−0.000	±0.002
GAL3$^+$	−	−0.726	±0.007
KlacGAL1$^+$	+ (s.)	−0.482	±0.017
gal1-ΔSA	−	−0.639	±0.007
gal3-∇SA	+ (v.s.)	−0.623	±0.005
klacgal1-ΔSA	−	−0.697	±0.010
gal1-atg1ctg	−	−0.624	±0.007
gal1-Δ	−	−0.724	±0.007

s., slow; v.s., very slow; s.d., standard deviation of three independent lines. Growth was scored manually in liquid culture using a single strain. Cultures contained 2% galactose as the sole carbon source. All data are normalized to the *GAL1*$^+$ control. *gal1-atg1ctg* (a mutation of the start codon) and *gal1-Δ* are both null mutations, but *gal1-Δ* causes an additional induction-dependent fitness defect, the basis of which is unknown. Note the poor performance of *KlacGAL1*$^+$ and *gal3-∇SA* relative to *GAL1*$^+$.

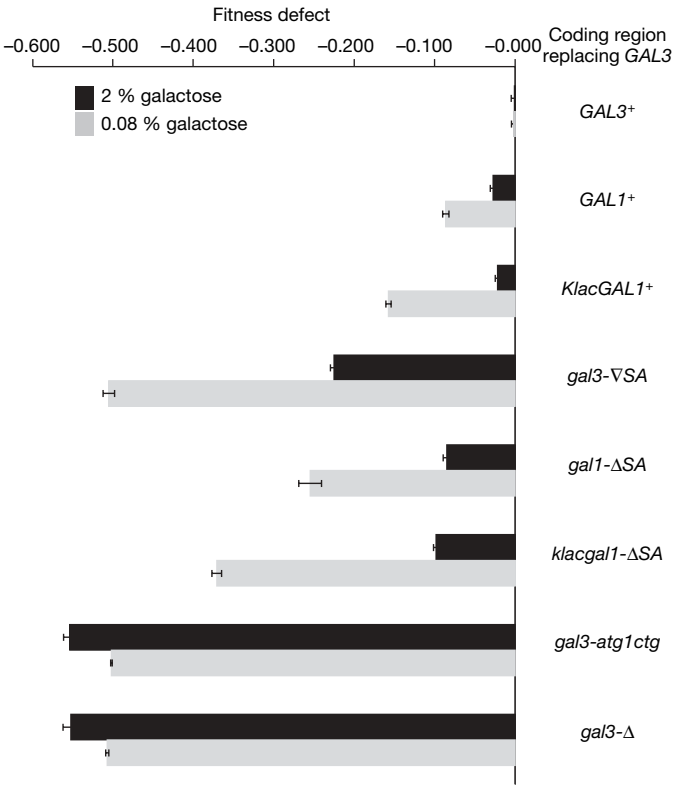


Figure 1 | *GAL3*$^+$ encodes the best co-inducer. The fitness effects of the targeted replacement of the coding region of *GAL3* under two different culture conditions are plotted as bar graphs. Error bars are standard deviations of three independent lines. All data are normalized to the *GAL3*$^+$ control. Note that *GAL3*$^+$ outperforms *GAL1*$^+$ and *KlacGAL1*$^+$, but the defect is moderate. Further note that the effect of the presence of the SA dipeptide depends on the identity of the coding region being tested.

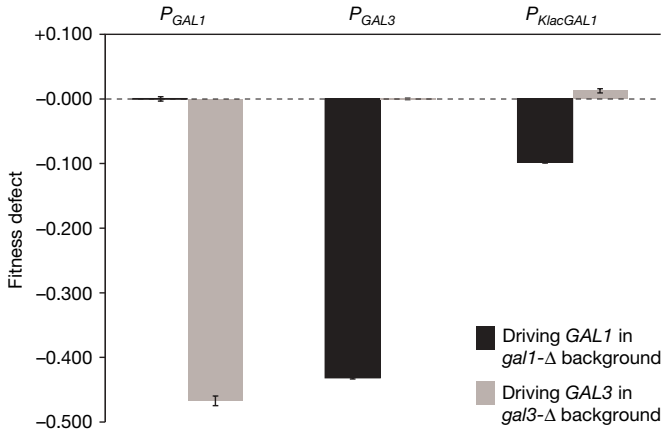


Figure 2 | Most adaptive divergence between *GAL1* and *GAL3* occurred in the promoters. Fitness effects of promoter replacements are shown as bar graphs. Black, competitions between strains with the stated promoter driving the expression of the *GAL1* coding region in a *gal1*-Δ background and a BFP-labelled P_{GAL1} strain in 2% galactose; data normalized to the P_{GAL1} strain. Grey, competitions between strains with the stated promoter driving the expression of the *GAL3* coding region in a *gal3*-Δ background and a BFP-labelled strain in 0.08% galactose; data normalized to the P_{GAL3} strain. Error bars are standard deviations of three independent lines. Note that P_{GAL1} and P_{GAL3} are both inferior at performing the function of the other promoter, and $P_{KlacGAL1}$ is not optimized for driving *GAL1* expression.

efficiently repressed in non-inducing conditions than P_{GAL3} (Table 2)[24,25]. When the co-inducer is driven by this tightly regulated promoter, basal expression is apparently too low to allow for efficient induction of the pathway when cells are presented with galactose. Quantitative comparisons of the fitness defects of these strains to the defects of strains harbouring null alleles suggests that the regulated expression of these two paralogues was almost completely subdivided between a loosely regulated promoter providing the basal expression necessary for pathway induction and a tightly regulated promoter driving high levels of the enzyme on induction.

Analysis of $P_{KlacGAL1}$ in *S. cerevisiae* provides additional support for the view that P_{GAL1} and P_{GAL3} were subfunctionalized from an ancestral bifunctional state. $P_{KlacGAL1}$ drives strong basal expression and performs well in the role of driving expression of the co-inducer Gal3, as measured by both fitness (Fig. 2) and mRNA expression (Table 2)[32,33]. Although it induces much more strongly than P_{GAL3}, this expression level still falls well short of the more optimal P_{GAL1} promoter (Table 2). This approximately 50% reduction in the level of *GAL1* induction driven by $P_{KlacGAL1}$ reduces organismal fitness by about 10% (Fig. 2), suggesting that P_{GAL1} has acquired adaptive mutations that allow it to induce more fully in the presence of galactose. We therefore wondered whether the demands on $P_{KlacGAL1}$ to serve as a bifunctional promoter driving the expression of a protein performing two distinct functions, one regulatory and one enzymatic, might disallow concurrent optimization of the expression of each function.

Adaptive conflict in phasing of binding sites. One striking difference between P_{GAL1} and $P_{KlacGAL1}$ is the spacing or helical phasing of binding sites for the transcriptional activator Gal4 (refs 16, 17, 24, 25, 33). *S. cerevisiae* and all closely related *Saccharomyces sensu stricto* species possessing P_{GAL1} have a core of three almost immediately adjacent Gal4 binding sites[37], which places the three bound Gal4 dimers on approximately the same side of the double helix. In contrast, the three core Gal4 binding sites of $P_{KlacGAL1}$ are spaced approximately half a helical turn apart such that the central Gal4 dimer is bound on the opposite side of the double helix (Fig. 3a)[33]. Previous analysis has demonstrated that Gal4 dimers that are nearby and on the same helical phase are cooperatively bound and more efficiently inhibited by the co-repressor Gal80 than either single binding sites or sites bound by Gal4 dimers on opposite helical phases[38].

To test whether the helical phasing of Gal4 binding sites might be responsible for some of the difference in expression between P_{GAL1} and $P_{KlacGAL1}$, we altered the spacing of the Gal4 binding sites in $P_{KlacGAL1}$ such that they were in phase, like in P_{GAL1}. Indeed, this alteration decreased non-induced $P_{KlacGAL1}$ expression by about 27% ($P < 10^{-4}$; presumably because of more efficient binding by Gal80 (ref. 38)) but also increased the induced expression by about 13% ($P < 10^{-2}$; perhaps owing to cooperative binding of co-activators by Gal4) (Fig. 3b; Supplementary Fig. 1b). Furthermore, these Gal4 binding site changes in $P_{KlacGAL1}$ increased fitness when driving expression of the enzyme (by about 1% at higher galactose concentrations; $P < 10^{-4}$), but they decreased fitness when driving

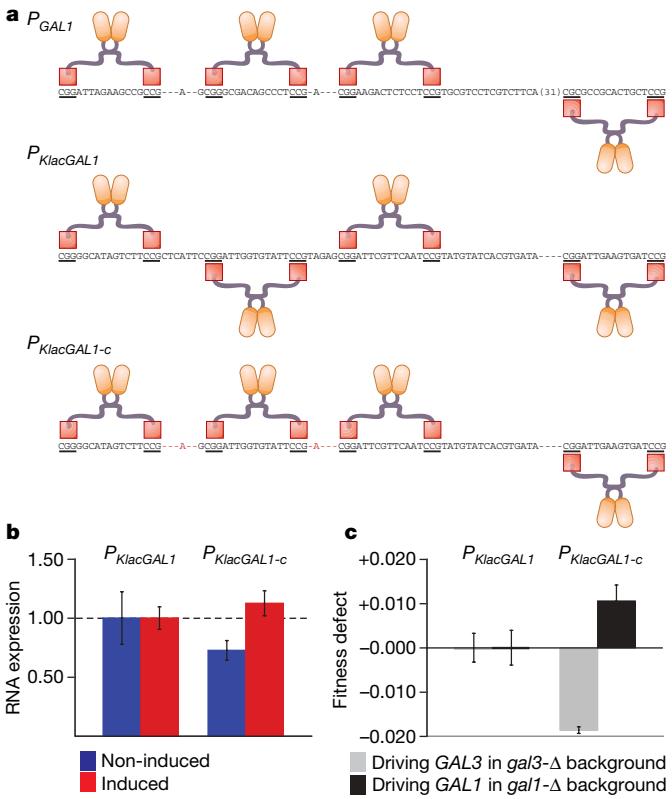


Figure 3 | Adaptive conflict in the configuration of Gal4 binding sites. **a**, Schematic of P_{GAL1}, $P_{KlacGAL1}$ and the mutant (nucleotide changes in red) $P_{KlacGAL1\text{-}c}$. **b**, *GAL1* mRNA expression normalized to $P_{KlacGAL1}$. Error bars are standard deviations of 12 replicates. **c**, Fitness effects of promoter replacement normalized to $P_{KlacGAL1}$ strains. Grey, competitions between a *gal3*-Δ strain with the stated promoter driving *GAL3* and a BFP-labelled strain in 0.08% galactose. Black, competitions between a *gal1*-Δ strain with stated promoter driving *GAL1* and a BFP-labelled P_{GAL1} strain in 2% galactose. Error bars are standard deviations of 3 independent lines and 12 cultures (4 each of three independent lines), respectively. Note that altering the phasing of $P_{KlacGAL1}$ to mimic P_{GAL1} improves fitness when driving *GAL1* as a result of an increase in the induced expression, but impairs fitness when driving *GAL3* owing to a decrease in the non-induced expression.

Table 2 | Normalized expression of promoters driving *GAL1*

Promoter	Non-induced	s.d.	Induced	s.d.
P_{GAL1}	1.00	±0.25	1.00	±0.13
P_{GAL3}	2.12	±0.38	0.13	±0.02
$P_{KlacGAL1}$	9.95	±1.64	0.54	±0.09

s.d., standard deviation of nine replicates. *GAL1* mRNA levels (measured by quantitative RT–PCR) driven from each promoter were normalized to *ACT1* mRNA levels and to the experimental value from P_{GAL1} for each condition. P_{GAL1} induces to ~1,000-fold above non-induced conditions[24,25].

expression of the co-inducer (by about 2% at lower galactose concentrations; $P < 10^{-2}$) (Fig. 3c; Supplementary Fig. 1c). Thus, alteration of binding site configuration caused opposing effects on the promoter's ability to perform its two distinct functions. These results suggest that an adaptive conflict was resolved by subdividing the transcriptional regulation of each function between two specialist genes during the evolution of the lineage that gave rise to *S. cerevisiae*.

The *K. lactis GAL1* promoter is not optimized. These results imply that some mutations that might have improved the performance of the promoter of the ancestral bifunctional *GAL1* gene were disfavoured before the gene duplication. We tested this idea further by examining the fitness impact of various promoters by placing them upstream of the bifunctional *K. lactis GAL1* gene in *K. lactis*.

We found that P_{GAL1} of *S. cerevisiae* actually outperformed $P_{KlacGAL1}$ at high galactose concentrations in *K. lactis* (Supplementary Tables 2–4). Similarly, $P_{KlacGAL1\text{-}\sigma}$ in which the Gal4 binding sites are placed on the same side of the helix, also outperformed $P_{KlacGAL1}$ at high galactose and lactose concentrations in *K. lactis*. Lactose is a common substrate for *K. lactis*, so it seems likely that *K. lactis* could benefit from *cis*-phasing of Gal4 binding sites in $P_{KlacGAL1}$ under some conditions. Yet, this configuration failed to evolve in *K. lactis*. The data from the fitness experiments in *S. cerevisiae* indicate that constraints imposed by the co-induction function of *KlacGAL1* may explain the suboptimal configuration of $P_{KlacGAL1}$.

The path to specialization. We have dissected the evolution and divergence of the paralogues *GAL1* and *GAL3* from a bifunctional ancestral gene by using precise genetic manipulation and a sensitive direct competition assay. In *S. cerevisiae*, these genes are nearly completely subfunctionalized, and both genes are required for efficient growth when galactose is the sole carbon source. Subfunctionalization of the coding regions was asymmetric: Gal3 completely lost galactokinase activity, whereas Gal1 retained substantial co-induction capability. Subfunctionalization of the promoters was nearly complete with P_{GAL1} and P_{GAL3} each unable to effectively perform the function of the other regulatory element. Comparison with $P_{KlacGAL1}$ suggests that P_{GAL1} has evolved tighter regulation and become more inducible. We showed that about 10% of this change (with respect to both fitness and gene expression level) was due to the

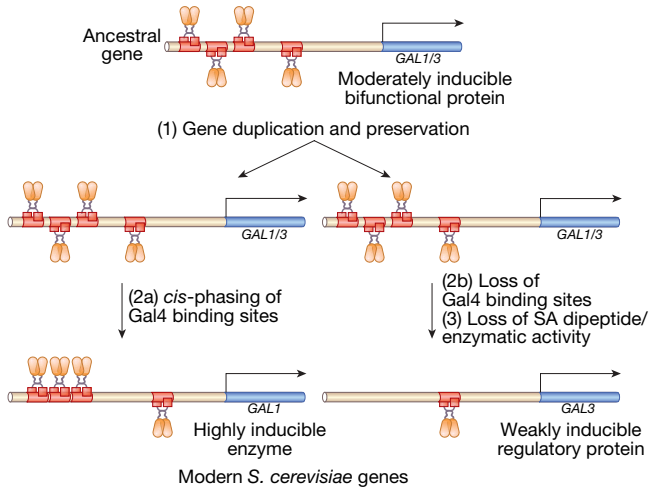


Figure 4 | Model of the evolution of the GAL genetic switch. Step wise subfunctionalization of the ancestral bifunctional *GAL1/3* gene. (1) Whole-genome duplication created two identical copies[19–23], which were probably preserved by DDC[11]. (2) The loss of Gal4 binding sites in P_{GAL3} and the evolution of *cis*-phasing of Gal4 binding sites in P_{Gal1} occurred in the lineage leading to the common ancestor of all *Saccharomyces sensu stricto*. *cis*-phasing of Gal4 binding sites helped optimize *GAL1* expression in a manner inconsistent with efficient co-induction in *S. cerevisiae*, resolving the putative adaptive conflict. (3) the loss of the SA dipeptide and the final loss of Gal3's enzymatic activity occurred within the *sensu stricto* clade[22].

helical phasing of Gal4 binding sites, although other important changes must also have occurred. The altered helical phasing of P_{GAL1} also decreased non-induced gene expression and conferred a fitness penalty on the promoter's ability to perform the co-induction function in *S. cerevisiae*. These findings suggest a scenario for the evolution of this genetic switch and the ultimate resolution of the underlying adaptive conflict (Fig. 4).

After the whole-genome duplication, *S. cerevisiae GAL1* and *GAL3* were integrated into a more complex and, in some ways, more optimal genetic pathway. Consistent with the DDC model, degeneration of complementary functions has been the major consequence. The most striking degeneration is the complete loss of galactokinase activity by Gal3, in which the final inactivating event (the loss of the SA dipeptide) occurred recently within the *sensu stricto* clade[22]. Our data demonstrate that the loss of enzymatic activity was not inherently adaptive but may have simply been the final step in a gradual degeneration that took nearly 100 million years to complete.

GAL3 may or may not have acquired any adaptive mutations of its own, but its retention as the co-inducer of the pathway allowed *GAL1* to evolve a more specialized regulatory profile (Fig. 4). We have demonstrated that the configuration of Gal4 binding sites upstream of *GAL1* is an important adaptive component of this regulatory logic that allows for stronger induction. However, this configuration is ill-suited to driving expression of the co-inducer in *S. cerevisiae* such that it is only adaptive when another co-inducer is present in the form of *GAL3*. We propose that these previously disfavoured mutations were adaptive for *GAL1* after the duplication of the ancestral *GAL1/3* and that subdividing the enzymatic and regulatory functions allowed this specific adaptive conflict to be resolved in the lineage that gave rise to *S. cerevisiae* (Fig. 4).

Optimization of regulation after duplication. Our demonstration of the optimization of *GAL1* expression after duplication raises the general questions of: how common are adaptive conflicts within regulatory elements, and what mechanisms may resolve them? Examination of whole-genome expression profiles has suggested that duplicated genes do generally evolve divergent expression patterns[39–43]. Some of the best-documented cases of neofunctionalization of coding sequences have also involved changes in gene expression[3,5,44]. The widespread divergence of gene expression following duplication events raises the possibility that some of the ancestral *cis*-regulatory elements may also have harboured adaptive conflicts.

In addition to the conflicts presented by the control of a bifunctional protein, we can envision two additional circumstances where an adaptive conflict may exist in regulatory elements. First, there are well-documented cases of 'enhancer-sharing', such as in the *Hox* gene complex, where a single enhancer governs the expression of two adjacent genes[45]. This phenomenon also occurs on a genome-wide level between adjacent and nearby genes in yeast[46]. Second, an enhancer may be bifunctional and drive the expression of one gene in two spatial domains within a multicellular organism. Like $P_{GAL1/3}$, such regulatory elements have the potential for adaptive conflict when the optimization of the expression of two different genes or of one gene in two different places may be in conflict. In such cases, duplication of regulatory sequences (or the entire gene) might allow for the evolution of novel gene expression profiles through previously forbidden mutations.

METHODS SUMMARY

Fluorescently labelled yeast strains were engineered to be genetically identical, except for tested mutations, and were competed against a standard strain. GFP/non-GFP ratios of live cells were determined by flow cytometry, and malthusian selection coefficients were calculated[47]. mRNA levels were monitored by quantitative PCR with reverse transcription (RT–PCR). See Methods for details.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 June; accepted 8 August 2007.

1. Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).
2. Taylor, J. S. & Raes, J. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38,** 615–643 (2004).
3. Irwin, D. M., Prager, E. M. & Wilson, A. C. Evolutionary genetics of ruminant lysozymes. *Anim. Genet.* **23,** 193–202 (1992).
4. Yokoyama, S. Molecular evolution of color vision in vertebrates. *Gene* **300,** 69–78 (2002).
5. Zhang, J., Zhang, Y. P. & Rosenberg, H. F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genet.* **30,** 411–415 (2002).
6. Prince, V. E. & Pickett, F. B. Splitting pairs: the diverging fates of duplicated genes. *Nature Rev. Genet.* **3,** 827–837 (2002).
7. van Hoof, A. Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171,** 1455–1461 (2005).
8. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440,** 341–345 (2006).
9. Schilke, B. *et al.* Evolution of mitochondrial chaperones utilized in Fe–S cluster biogenesis. *Curr. Biol.* **16,** 1660–1665 (2006).
10. Piatigorsky, J. *Gene Sharing and Evolution: The Diversity of Protein Functions* (Harvard Univ. Press, Cambridge, Massachusetts, 2007).
11. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–1545 (1999).
12. Tumpel, S., Cambronero, F., Wiedemann, L. M. & Krumlauf, R. Evolution of *cis* elements in the differential expression of two *Hoxa2* coparalogous genes in pufferfish (*Takifugu rubripes*). *Proc. Natl Acad. Sci. USA* **103,** 5419–5424 (2006).
13. Piatigorsky, J. & Wistow, G. The recruitment of crystallins: new functions precede gene duplication. *Science* **252,** 1078–1079 (1991).
14. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* **256,** 119–124 (1994).
15. Hughes, A. L. *Adaptive Evolution of Genes and Genomes* (Oxford Univ. Press, Oxford, 1999).
16. Johnston, M. A model fungal gene regulatory mechanism: the *GAL* genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **51,** 458–476 (1987).
17. Bhat, P. J. & Murthy, T. V. Transcriptional control of the GAL/MEL regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. *Mol. Microbiol.* **40,** 1059–1066 (2001).
18. Ptashne, M. & Gann, A. *Genes and Signals* (Cold Spring Harbor Laboratory Press, Woodbury, New York, 2001).
19. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387,** 708–713 (1997).
20. Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304,** 304–307 (2004).
21. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428,** 617–624 (2004).
22. Hittinger, C. T., Rokas, A. & Carroll, S. B. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl Acad. Sci. USA* **101,** 14144–14149 (2004).
23. Scannell, D. R. *et al.* Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl Acad. Sci. USA* **104,** 8397–8402 (2007).
24. Johnston, M. & Davis, R. W. Sequences that regulate the divergent *GAL1–GAL10* promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4,** 1440–1448 (1984).
25. West, R. W. Jr, Yocum, R. R. & Ptashne, M. *Saccharomyces cerevisiae GAL1–GAL10* divergent promoter region: location and function of the upstream activating sequence *UASG*. *Mol. Cell. Biol.* **4,** 2467–2478 (1984).
26. Bajwa, W., Torchia, T. E. & Hopper, J. E. Yeast regulatory gene *GAL3*: carbon regulation; *UASGal* elements in common with *GAL1*, *GAL2*, *GAL7*, *GAL10*, *GAL80*, and *MEL1*; encoded protein strikingly similar to yeast and *Escherichia coli* galactokinases. *Mol. Cell. Biol.* **8,** 3439–3447 (1988).
27. Zenke, F. T. *et al.* Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p. *Science* **272,** 1662–1665 (1996).
28. Yano, K. & Fukasawa, T. Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **94,** 1721–1726 (1997).
29. Platt, A. & Reece, R. J. The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex. *EMBO J.* **17,** 4086–4091 (1998).
30. Peng, G. & Hopper, J. E. Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. *Proc. Natl Acad. Sci. USA* **99,** 8548–8553 (2002).
31. Meyer, J., Walker-Jonah, A. & Hollenberg, C. P. Galactokinase encoded by *GAL1* is a bifunctional protein required for induction of the *GAL* genes in *Kluyveromyces lactis* and is able to suppress the *gal3* phenotype in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11,** 5454–5461 (1991).
32. Rubio-Texeira, M. A comparative analysis of the *GAL* genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. *FEMS Yeast Res.* **5,** 1115–1128 (2005).
33. Webster, T. D. & Dickson, R. C. The organization and transcription of the galactose gene cluster of *Kluyveromyces lactis*. *Nucleic Acids Res.* **16,** 8011–8028 (1988).
34. Thoden, J. B., Sellick, C. A., Timson, D. J., Reece, R. J. & Holden, H. M. Molecular structure of *Saccharomyces cerevisiae* Gal1p, a bifunctional galactokinase and transcriptional inducer. *J. Biol. Chem.* **280,** 36905–36911 (2005).
35. Platt, A., Ross, H. C., Hankin, S. & Reece, R. J. The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. *Proc. Natl Acad. Sci. USA* **97,** 3154–3159 (2000).
36. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evol. Int. J. Org. Evol.* **59,** 1165–1174 (2005).
37. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254 (2003).
38. Melcher, K. & Xu, H. E. Gal80–Gal80 interaction on adjacent Gal4p binding sites is required for complete *GAL* gene repression. *EMBO J.* **20,** 841–851 (2001).
39. Wagner, A. Asymmetric functional divergence of duplicate genes in yeast. *Mol. Biol. Evol.* **19,** 1760–1768 (2002).
40. Makova, K. D. & Li, W. H. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13,** 1638–1645 (2003).
41. Gu, Z., Rifkin, S. A., White, K. P. & Li, W. H. Duplicate genes increase gene expression diversity within and between species. *Nature Genet.* **36,** 577–579 (2004).
42. Huminiecki, L. & Wolfe, K. H. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome Res.* **14,** 1870–1879 (2004).
43. Li, W. H., Yang, J. & Gu, X. Expression divergence between duplicate genes. *Trends Genet.* **21,** 602–607 (2005).
44. Goodman, M., Moore, G. W. & Matsuda, G. Darwinian evolution in the genealogy of haemoglobin. *Nature* **253,** 603–608 (1975).
45. Gould, A., Morrison, A., Sproat, G., White, R. A. & Krumlauf, R. Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping *Hox* expression patterns. *Genes Dev.* **11,** 900–913 (1997).
46. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26,** 183–186 (2000).
47. Hartl, D. L. & Clark, A. G. in *Principles of Population Genetics* 216 (Sinauer Associates, Sunderland, Massachusetts, 1997).
48. Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–132 (1998).
49. Kooistra, R., Hooykaas, P. J. & Steensma, H. Y. Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* **21,** 781–792 (2004).
50. Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24,** 2519–2524 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. L. Williams for sharing unpublished data, reagents and the competition assay; J. E. Selegue for technical support; M. Johnston and Washington University in St Louis for access to laboratory space and equipment; Collection de Levures d'Intérêt Biotechnologique for *K. lactis* strain CLIB210; H. Y. Steensma for the NHEJ-deficient *K. lactis* strain and transformation protocol; J. H. Hegemann for the *kanMX* cassette; B. L. Williams, P. M. van Wynsberghe and Carroll laboratory members for technical advice; L. M. Olds for artwork; and B. L. Williams, B. Prud'homme and M. Johnston for critical reading of the manuscript. The Howard Hughes Medical Institute supported this work through an investigatorship (S.B.C.) and predoctoral fellowship (C.T.H.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.B.C. (sbcarrol@wisc.edu) and C.T.H. (cthittinger@wustl.edu).

METHODS

Strain construction. *MAT**a** ura3*-Δ *lys2*-Δ BY4724 strains of *S. cerevisiae*[48] were genetically engineered to contain only the desired mutations by standard aptamer-*URA3*-based selection and counter-selection. GFP was expressed from a P_{TDH3}-*yEGFP*-T_{CYC1} or P_{TDH3}-*yEBFP*-T_{CYC1} construct integrated into the genome between base pairs 199,270 and 199,271 of chromosome I. GFP- and BFP-labelled strains differed by a single base pair that encoded a single substitution in the GFP/BFP chromophore (Y66F) and were equally fit in competition assays. Coding regions were replaced from start codon to stop codon in the GFP-labelled strain, including control constructs that were engineered back to wild type. Promoter manipulations were carried out by inserting the desired promoter (defined as the entire upstream intergenic region) upstream of the desired coding region, with the entire 3′ intergenic region and no intervening sequence, between base pairs 199,230 and 199,231 of chromosome I.

*MAT**a** ku80*-Δ CBS2359 NHEJ-deficient strains of *K. lactis*[49] were genetically engineered to contain only the desired mutations. First, a *klacgal1*Δ::*kanMX* strain was constructed by transforming *K. lactis*[49] with a PCR product derived from a pUG6 template and with 90-mers that contained the *kanMX* amplification sites[50] fused to portions of the *K. lactis* genome necessary to replace the coding region of *KlacGAL1* with the *kanMX* cassette in the forward orientation from stop codon to start codon. A single kan^+ gal^- strain was transformed in parallel with PCR products constructed with 90-mers designed to target each promoter fused to the complete coding and 3′ intergenic region of *KlacGAL1* to the *URA3* locus. These products replaced the coding region of *URA3* in the reverse orientation from stop codon to start codon (*MAT**a** ku80*-Δ *klacgal1*Δ::*kanMX ura3*Δ::P_X-*KlacGAL1* CBS2359). Putative transformants were replica-plated from galactose plates to YPD plates containing 5-fluoro-orotic acid and screened by PCR. The *K. lactis* GFP-labelled competitor also contained the GFP construct from the *S. cerevisiae* GFP strains fused immediately downstream of the inserted 3′ UTR of *KlacGAL1* in the forward orientation (*MAT**a** ku80*-Δ *gal1*Δ::*kanMX ura3*Δ::$P_{KlacGAL1}$-*KlacGAL1*::P_{TDH3}-*yEGFP*-T_{CYC1} CBS2359). The *K. lactis* GFP-labelled strain used contained 14 consecutive A's instead of 15 A's in the region covered by the forward targeting primer at the 5′ end of the promoter. This strain produced data similar to another *K. lactis* GFP-labelled strain that had 15 consecutive A's but was missing a base pair in the former promoter of *URA3* in the region covered by the reverse targeting primer.

All experimental strains from both species were engineered in triplicate, and each manipulated region was sequence-verified. All primer sequences are available on request.

Competition assays. For each *S. cerevisiae* competition experiment, frozen stocks were plated to YPD. Individual colonies were cultured in autoclaved SC (made according to the manufacturer's instructions with Complete Supplement Mixture (Qbiogene) and Yeast Nitrogen Base without Amino Acids (Amresco)) containing 2% autoclaved raffinose (non-inducing) for 2 days at 30 °C on a rotor wheel. Saturated cultures of each strain were added to aliquots from a single BFP culture. These approximately 50% GFP mixtures were used to inoculate fresh SC with 2% raffinose cultures, which were grown for 2 days. A portion of each culture was then used to inoculate SC with X% galactose at a known concentration, which was varied so that the number of generations was between eight and ten. A portion of this culture was also used to determine the proportion of GFP cells at the start of the experiment (GFPstart) by counting 2×10^5 live cells (those not staining with propidium iodide) and scoring for the presence or absence of GFP, using flow cytometry. After 2 days in SC with X% galactose, a portion of this culture was also used to determine the proportion of GFP cells at the end of the experiment (GFPend) in the same manner. The exact number of generations (t) for each experiment was determined using values from cell counts for a standard GFP versus BFP competition and the known dilution factors. Malthusian selection coefficients (m) were then calculated as $m = \ln(10^{\log[(\mathrm{GFPend}/\mathrm{BFPend})/(\mathrm{GFPstart}/\mathrm{BFPstart})]/t})$ derived from equation (6.4) in ref. 47: $A(t)/B(t) = A(0)/B(0)\ e^{mt}$. For a given experiment performed in parallel against aliquots of the same culture of the stated BFP competitor strain, selection coefficients were normalized by subtracting the measured selection coefficient of the stated GFP control strain. Experiments were carried out in similar fashion in *K. lactis*, except that filter-sterilized carbon sources were used, sucrose was used as the non-inducing carbon source, and the GFP-labelled strain was the competitor strain. For *K. lactis*, at least 10^5 live cells were counted, except for in 0.08% and 0.2% galactose where at least 10^4 cells were counted.

Quantification of mRNA expression. Total RNA was obtained from early log phase cells grown in SC with 2% raffinose before (non-induced) or after (induced) a 60-min induction, with galactose added to a final concentration of 2%. The strains used were the same used in the competition assays with the tested promoter driving the expression of *GAL1* in a *gal1-Δ* background. DNase-treated total RNA (5–25 ng) was used in custom probe-based Taqman One-Step RT–PCR assays using a 7900HT Fast Real-Time PCR System (Applied Biosystems) according to the manufacturer's instructions. The experimental primer/probe set monitored *GAL1* mRNA expression from various promoters, whereas *ACT1* mRNA was a control for RNA quantity in separate reactions. *GAL1* primers were CGTTTATTATGCCAGATATCACAACA (forward; 1,400 nM) and CCGTTCGATGCCGGATT (reverse; 1,400 nM) with a 5′ 6-FAM- and 3′ TAMRA-labelled probe (200 nM), TTCCACACCCTGGAACGGCGAT (Integrated DNA Technologies). *ACT1* primers were GTGATGTCGATGTCCGTAAGGAA (forward; 1,400 nM) and TCTGGAGGAGCAATGATCTTGAC (reverse; 1,400 nM) with a 5′ 6-FAM- and 3′ TAMRA-labelled probe, ATCACCGCTTTGGCTCCATCTTCCAT (100 nM or 200 nM). Twenty microlitre reaction conditions were 30 min at 48 °C, 10 min at 95 °C, and 40 cycles of alternating 15 s at 95 °C and 60 s at 60 °C. After applying the recommended background correction from cycles 3 through 15, threshold cycle (C_t) values were obtained with a threshold of 0.2 using the manufacturer's software. Failed reactions, whose C_t was undetermined or greater than one C_t from the next nearest reading, were excluded. Each biological replicate was supported by at least two RT–PCR replicates. C_t values were converted to relative quantities, normalized to *ACT1* mRNA expression and normalized to the stated promoter.

Statistics. All P values given are from one-tailed non-parametric tests performed using Mstat version 4.01 (http://mcardle.oncology.wisc.edu/mstat/). Gene expression data from several experiments were pooled before analysis with Wilcoxon rank sum tests. For *S. cerevisiae*, fitness values varied (in magnitude, but not relative rankings) with the number of generations and batch of culture media, especially for highly negative values. Therefore, fitness data reported are for a representative experiment employing each of the three independently engineered strains in parallel, but all data were incorporated into statistical tests using Lehman's test, a multiple experiment permutation test based on the Wilcoxon rank sum test. Wilcoxon rank sum tests were used on the *K. lactis* fitness data.

ARTICLES

Tumour invasion and metastasis initiated by microRNA-10b in breast cancer

Li Ma[1], Julie Teruya-Feldstein[2] & Robert A. Weinberg[1]

MicroRNAs have been implicated in regulating diverse cellular pathways. Although there is emerging evidence that some microRNAs can function as oncogenes or tumour suppressors, the role of microRNAs in mediating cancer metastasis remains unexplored. Here we show, using a combination of mouse and human cells, that microRNA-10b (miR-10b) is highly expressed in metastatic breast cancer cells and positively regulates cell migration and invasion. Overexpression of miR-10b in otherwise non-metastatic breast tumours initiates robust invasion and metastasis. Expression of miR-10b is induced by the transcription factor Twist, which binds directly to the putative promoter of *mir-10b* (*MIRN10B*). The miR-10b induced by Twist proceeds to inhibit translation of the messenger RNA encoding homeobox D10, resulting in increased expression of a well-characterized pro-metastatic gene, *RHOC*. Significantly, the level of miR-10b expression in primary breast carcinomas correlates with clinical progression. These findings suggest the workings of an undescribed regulatory pathway, in which a pleiotropic transcription factor induces expression of a specific microRNA, which suppresses its direct target and in turn activates another pro-metastatic gene, leading to tumour cell invasion and metastasis.

Metastasis is a complex, multi-step process by which primary tumour cells invade adjacent tissue, enter the systemic circulation (intravasate), translocate through the vasculature, arrest in distant capillaries, extravasate into the surrounding tissue parenchyma, and finally proliferate from microscopic growths (micrometastases) into macroscopic secondary tumours[1]. In recent years, studies have been carried out to investigate the genes and gene products that drive the metastatic process. For instance, work in a number of laboratories has revealed several transcription factors that can program many of the cell-biological changes needed to execute the initial steps of the invasion–metastasis cascade[2–8].

Lately, it has become evident that, in addition to alterations in protein-encoding genes, abnormalities in non-coding genes can also contribute to cancer pathogenesis[9,10]. In particular, a class of small cellular RNAs, termed microRNAs (miRNAs), acting as agents of the RNA interference pathway, can lead to silencing of their cognate target genes, doing so either by cleaving mRNA molecules or by inhibiting their translation[11]. Indeed, miRNAs have been implicated in the regulation of a variety of cellular processes, including apoptosis[12], haematopoietic differentiation[13], metabolism[14], skin morphogenesis[15] and neural development[16].

More than 50% of annotated human miRNA genes are located in fragile chromosomal regions that are susceptible to amplification, deletion, or translocation during the course of tumour development[17]. Moreover, recent evidence indicates that some miRNAs can function either as oncogenes or tumour suppressors[10,18,19], and expression profiling analyses have revealed characteristic miRNA signatures in certain human cancers[9,20,21]. However, the precise parts played by the expressed miRNAs in specific steps of malignant progression, including metastasis, are still unknown. For these reasons, we undertook to associate specific miRNAs with specific stages of malignant progression, with the hope that such associations might provide insights into the causal mechanisms of cancer cell invasion and metastasis.

miR-10b is highly expressed in metastatic breast cancer cells

To identify miRNAs that regulate breast cancer metastasis, we selected candidate miRNAs on the basis of previously reported miRNA microarray profiling; these analyses had identified 29 miRNAs that are differentially expressed between primary breast carcinomas and normal mammary tissue, without regard to the eventual metastatic progression of these tumours[22]. We investigated the expression of these candidate miRNAs in a series of human mammary epithelial cells and tumour cell lines (Supplementary Table 1). Out of a total of eight selected miRNAs, three (miR-155, miR-9 and miR-10b) were found to be markedly upregulated in breast cancer cells when compared with either primary human mammary epithelial cells (HMECs) or with the spontaneously immortalized MCF-10A cells (Fig. 1a and Supplementary Table 1). Unlike miR-155 and miR-9, the expression of which was not specific to metastatic cells (Fig. 1a, b), miR-10b was highly expressed only in metastatic cancer cells. For example, the expression level of miR-10b was 50-fold higher in cells of the MDA-MB-231 line, which are capable of metastasizing, than in cells of the MCF-7 human breast cancer line, which have little if any metastatic powers (Fig. 1c). This correlation indicated that miR-10b might well have a causal role in breast cancer metastasis.

miR-10b positively regulates cell migration and invasion *in vitro*

We first performed *in vitro* loss-of-function analyses by silencing the miRNAs with antisense oligonucleotides[23]. We assessed the level of miRNA silencing by a reporter assay, in which the predicted miRNA binding site was cloned into the 3′ untranslated region (UTR) of a luciferase reporter gene[24]. We found that transfection of the antisense inhibitor for miR-9 or miR-10b in MDA-MB-231 cells caused a two- to threefold increase in the luciferase activity (Supplementary Fig. 1a, b), suggesting that each of the transfected antisense RNAs achieved a greater than 50% inhibition of the actions of its cognate miRNA. Although neither of these two antisense RNAs affected the motility of MDA-MB-231 cells (Fig. 1d), silencing of miR-10b led to a more

[1]Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [2]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

than tenfold reduction in the invasive properties of these cells, as gauged by an *in vitro* invasion assay in which miR-9 inhibition had only a marginal effect (Fig. 1e). This reduction was not due to impairment of cell viability (Supplementary Fig. 1c). Taken together, these observations suggested that miR-10b function is required for *in vitro* invasiveness but not for viability or motility of these metastatic cells.

To determine whether miR-10b overexpression would increase the basal levels of cell migration or invasion, we cloned the genomic sequence of the human *mir-10b* gene into a green fluorescent protein (GFP)-expressing, murine stem-cell retrovirus (MSCV)-derived vector[13]. We then used the resulting vector to express miR-10b in immortalized HMECs[25], and in the SUM149 cell line, a line of non-metastatic human breast cancer cells[26,27]. The miR-10b expression level was gauged by PCR with reverse transcription (RT–PCR; Fig. 1f).

In both cell lines, ectopic expression of miR-10b had no effect on their proliferation *in vitro* (Supplementary Fig. 1d, e), but did result in a four- to sixfold increase in cell motility and invasiveness (Fig. 1g, h). These results indicated that overexpression of miR-10b is sufficient to promote both migration and invasion *in vitro*.

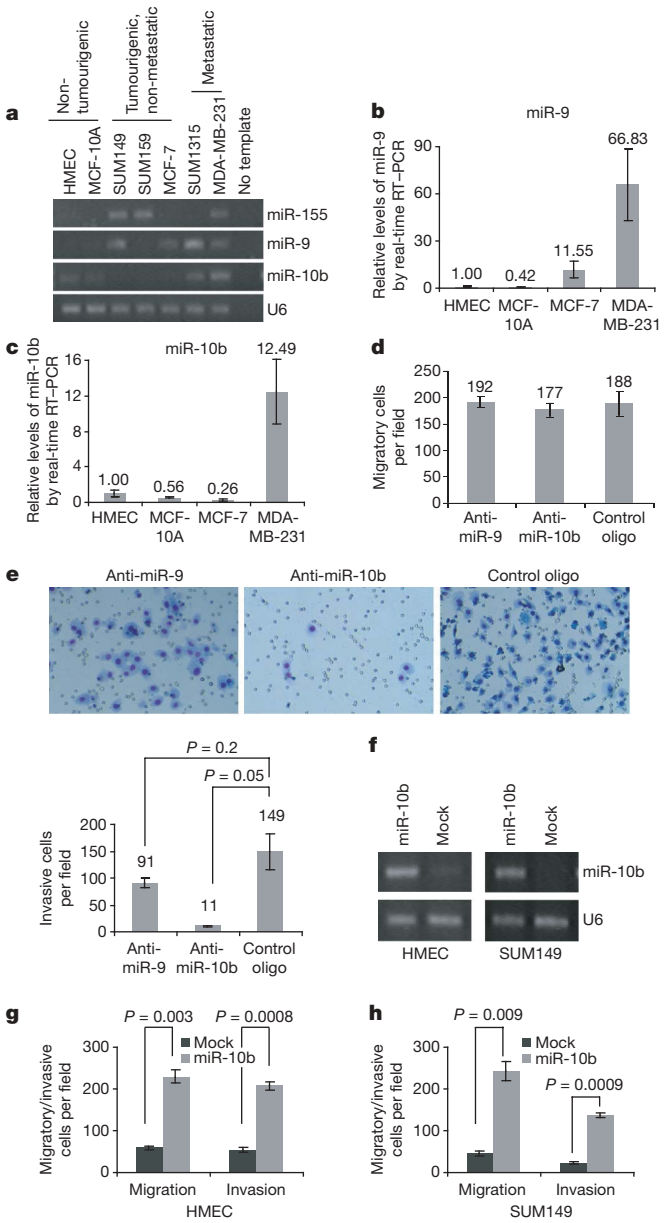


Figure 1 | miR-10b is highly expressed in metastatic breast cancer cells and positively regulates cell migration and invasion. **a**, RT–PCR of miR-155, miR-9 and miR-10b in a series of human mammary epithelial cells. **b**, **c**, Real-time RT–PCR of miR-9 (**b**) and miR-10b (**c**) in four different human mammary epithelial cells. **d**, **e**, Transwell migration assay (**d**) and Matrigel invasion assay (**e**) of MDA-MB-231 cells transfected with the inhibitor for miR-9 (anti-miR-9) or miR-10b (anti-miR-10b) (quantified below). Magnification in **e**, ×200. **f**, RT–PCR of miR-10b in HMECs and SUM149 cells infected with the miR-10b-expressing or empty vector. **g**, **h**, Transwell migration assay and Matrigel invasion assay of HMECs (**g**) and SUM149 cells (**h**) infected with the miR-10b-expressing or empty vector. A representative experiment is shown in triplicate along with s.e.m. in **b**–**e**, **g** and **h**.

miR-10b initiates tumour invasion *in vivo*

A key question was whether miR-10b could induce metastatic behaviours *in vivo*. To address this possibility, we overexpressed miR-10b in otherwise non-metastatic human breast cancer cells. To begin, we implanted miR-10b-transduced or mock-infected SUM149 cells into the mammary fat pads of NOD-SCID mice. The host mice displayed visible mammary tumours within two weeks after injection and became moribund at week 11 owing to primary tumour burden, at which point this experiment was terminated. Importantly, GFP expression was maintained in the tumour cells throughout this time period (Fig. 2a).

At 6 weeks post implantation, the control tumours and the miR-10b-overexpressing SUM149 tumours were of comparable size (Fig. 2b), indicating no effects of this miRNA on primary tumour formation. As anticipated, the control SUM149 tumours were strictly non-invasive, as shown by their confinement within fibrotic capsules (Fig. 2c, panels A, B). In stark contrast, the miR-10b-overexpressing SUM149 tumours displayed a massive desmoplastic reaction, with islands of epithelial cancer cells that had invaded the stroma (Fig. 2c, panels C, D). In addition, we observed apparent muscular and vascular invasion by the miR-10b-overexpressing tumour cells (Fig. 2c, panels E, F). Hence, ectopic expression of miR-10b could confer invasiveness on SUM149 cancer cells that were previously non-invasive *in vivo*.

To determine whether miR-10b expression in the primary tumours would also affect cell proliferation and tumour angiogenesis, we performed immunohistochemistry with the Ki-67 proliferation marker and the MECA-32 endothelial cell marker. We found that the distribution, but not the total number, of Ki-67$^+$ cells in the miR-10b-overexpressing SUM149 tumours was distinct from that seen in the control tumours: in the miR-10b-expressing tumours, large necrotic centres were apparent that lacked proliferating cells, whereas the invasion fronts were enriched in Ki-67$^+$ cells and were highly vascularized (Fig. 2d, panel E–H, and Fig. 2e, f); this contrasted with the appearance of the control tumours, which exhibited an essentially even distribution of Ki-67$^+$ cells (interspersed with Ki-67$^-$ cells) and poor vascularization (Fig. 2d, panel A–D, and Fig. 2e, f). Furthermore, in contrast to the control tumours, the vessels associated with the invasion fronts of the miR-10b-overexpressing tumours were seen not only in the stroma (peritumoural), but also within the epithelial tumour masses (intratumoural, Fig. 2g). Hence, the invasion fronts of miR-10b-overexpressing tumours exhibited very high levels of both cell proliferation and angiogenesis.

At later times, those mice injected with miR-10b-overexpressing cells carried larger tumour burdens ($P = 0.03$, Fig. 2b). We concluded that the enhanced tumour growth of these miR-10b-overexpressing cells was likely to be a consequence rather than a cause of their invasiveness, because (1) the substantial stromal invasion by the miR-10b-expressing cells preceded by several weeks the observed differences in overall tumour growth rates, and (2) the miR-10b-transduced cells and their control counterparts proliferated at similar rates *in vitro* (Supplementary Fig. 1d, e) and *in vivo* (as gauged by overall levels of Ki-67, Fig. 2d, e). Thus, we speculate that vascularization might be a rate-limiting determinant of overall tumour growth, and that increased invasiveness might afford the miR-10b-overexpressing cells better access to the vasculature.

miR-10b initiates distant metastasis

We asked whether expression of miR-10b would also result in distant metastasis. As early as week 6, haematoxylin and eosin staining revealed the presence of miR-10b-overexpressing SUM149 cells in the primary tumour-associated vessels (Fig. 2c, panel F). Immunohistochemical analyses further revealed significant numbers of hyperproliferative Ki-67$^+$ tumour cells in the lumina of some of the vessels associated with the miR-10b-overexpressing tumours (Fig. 3a). In contrast, no significant evidence of intravasating or intravasated cancer cells was found in the control tumours.

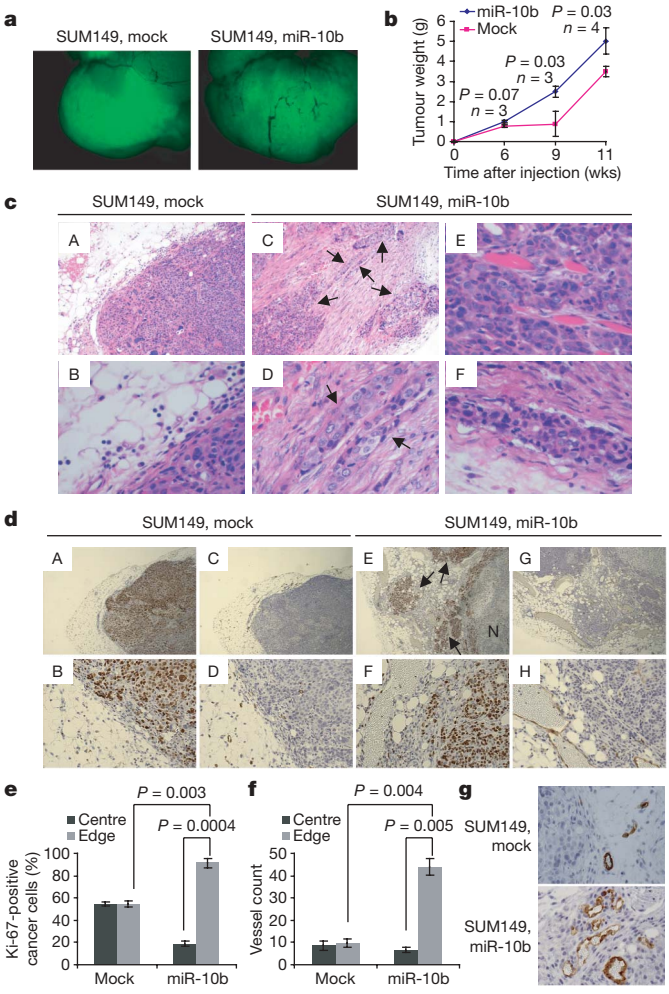


Figure 2 | miR-10b induces tumour invasion. **a**, GFP imaging of the SUM149 primary tumours at week 11 after orthotopic injection of NOD-SCID mice. Magnification, ×8. **b**, Growth curves of primary mammary tumours formed by SUM149 cells infected with the miR-10b-expressing or empty vector. Each data point represents the mean ± s.e.m. of 3–4 mice. **c**, Haematoxylin and eosin (H&E)-stained sections of primary mammary tumours formed by SUM149 cells infected with the miR-10b-expressing or empty vector, at week 6 after orthotopic transplantation. Arrows in panels C and D indicate areas of stromal invasion. Panels E and F demonstrate muscular invasion and vascular invasion, respectively. Magnification: panels A and C, ×100; panels B and D–F, ×400. **d**, Ki-67- (panels A, B, E, F) and MECA-32- (panels C, D, G, H) stained sections of primary mammary tumours formed by SUM149 cells infected with the miR-10b-expressing or empty vector, at week 6 after orthotopic transplantation. Arrows in panel E indicate areas of invasion. N, necrosis. Magnification: panels A, C, E and G, ×40; panels B, D, F and H, ×200. **e**, **f**, Quantification of Ki-67 staining (percentage of Ki-67$^+$ carcinoma cells among total carcinoma cells; **e**) and vessels (using MECA-32-stained sections; **f**) at the centre and the edge of the SUM149 tumours. $n = 3$ mice at 6 weeks after implantation. Error bars in **e** and **f** indicate s.e.m. **g**, Prominent intratumoural vessels are associated with the invasion front of miR-10b-overexpressing tumours, as demonstrated by MECA-32 staining of primary mammary tumours formed by SUM149 cells infected with the miR-10b-expressing or empty vector, at week 6 after orthotopic transplantation. Magnification, ×400.

We also examined other tissues for the presence of disseminated tumour cells. At week 6 after implantation, we observed occasional single GFP$^+$ cells in a few regions of the lung cryosections, indicating relatively early dissemination of the miR-10b-overexpressing cells from primary tumours; the lungs of mice that carried control tumours lacked such GFP$^+$ cells (Supplementary Fig. 2). At 9 weeks after implantation, the lungs from mice implanted with miR-10b-overexpressing SUM149 cells exhibited clusters of dense hyperchromatic cells that were positive for cytokeratins, as demonstrated by staining with AE1/AE3 (a cocktail of two distinct anti-cytokeratin monoclonal antibodies, Fig. 3b). On average, we found ~1 micrometastasis per 5-μm section (Fig. 3c, left panel). At week 11, there was a further increase in the number of such micrometastatic clusters (~4 micrometastases per section, Fig. 3c, right panel). In the lungs of hosts bearing control SUM149 tumours, however, the only AE1/AE3-positive cells were normal bronchial epithelial cells (Fig. 3b).

The ability of ectopically expressed miR-10b to elicit metastasis was also examined in the SUM159 cell line—a second line of human breast cancer cells. These cells are invasive but non-metastatic[27], allowing us to determine whether these cells too would respond to miR-10b by acquiring metastatic potential. In contrast to the SUM149 cells described above, control SUM159 cells exhibited a high level of motility and invasiveness *in vitro* (data not shown). Moreover, the xenograft mammary tumours formed by the control SUM159 cells displayed local invasion (data not shown). Importantly, no metastases were found in 10 mice that had been transplanted with mock-infected SUM159 cells (Fig. 3d, f). Strikingly, however, 8 out of 10 mice that had received orthotopic transplantation of miR-10b-overexpressing SUM159 cells exhibited numerous lung metastases, which were readily detectable both by GFP fluorescence and by histological analysis (Fig. 3d, f). Furthermore, 30% of these mice (3/10) developed macroscopic peritoneal metastases (0.5–2 cm in diameter, Fig. 3e, f). Taken together, these observations indicate that ectopic miR-10b expression can drive tumour invasion and metastasis in otherwise non-metastatic breast tumours, thereby acting as a potent pro-metastatic agent.

miR-10b is directly regulated by the transcription factor Twist

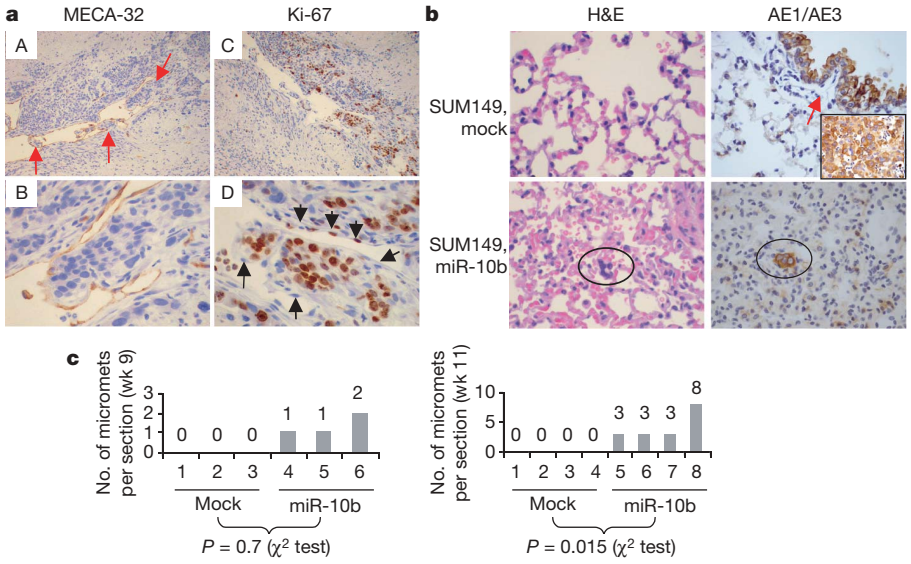
We undertook to determine how miR-10b expression is regulated. The observation that miR-10b is highly expressed in metastatic breast cancer cells indicated that the gene encoding this miRNA might be the target of certain transcription factors that are activated specifically in metastatic cells. Of note, recent studies have demonstrated that several transcription factors previously known as master regulators of embryogenesis, are highly expressed in metastatic cells and seem to have causal roles in tumour metastasis, ostensibly by inducing epithelial–mesenchymal transitions (EMTs) in cancer cells[2–8]; this process is thought to contribute to the invasiveness and dissemination of epithelial tumour cells[28]. These various observations led us to ask whether any of these EMT-inducing factors might function to activate miR-10b expression.

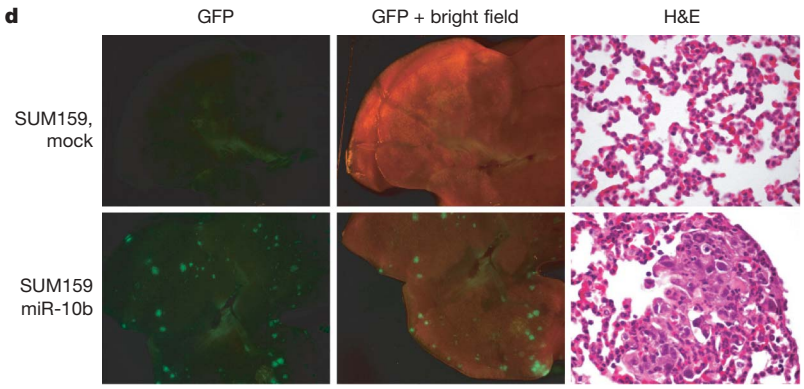
We first assessed miR-10b expression in the four lines of mouse mammary tumour cells that had been used in the identification of *Twist* as a metastasis-promoting gene[6]. We discovered that the level of miR-10b correlated with the known metastatic potentials of these cell types, with the lowest expression level in 67NR cells, which are unable to intravasate from the primary tumour, whereas the highest expression level was seen in 4T1 cells, which are capable of generating macroscopic metastases (Fig. 4a, b). This closely paralleled the expression pattern of Twist in these cell lines[6], indicating that miR-10b expression might well be upregulated by this transcription factor. We therefore expressed either Twist1 (also known as Twist) or, as a control, a second EMT-inducing transcription factor—snail (SNAI1)—in the non-tumourigenic, immortalized HMECs, which had been found to express a low level of miR-10b (Fig. 1f). In contrast

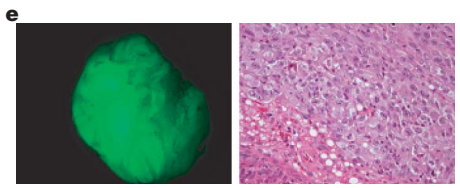
to the behaviour of SNAI1, which reduced miR-10b expression by 57%, ectopic expression of Twist1 led to a 4.5-fold increase in the level of this miRNA in these HMECs (Fig. 4c, d).

We next performed chromatin immunoprecipitation (ChIP) assays to determine whether TWIST1 controls miR-10b expression by binding directly to the *mir-10b* gene. This basic helix–loop–helix transcription factor has been shown to bind to E-box sequences (CANNTG) present in the genes that it regulates[29,30]. We examined the 4-kb genomic sequence upstream of the human *mir-10b* stem-loop and identified two conserved E-boxes, at −313 bp (E-box 1) and −2,422 bp (E-box 2), respectively (Fig. 4e). In addition, the putative promoter of human *mir-10b* spans between −111 bp and −460 bp[31], which encompasses E-box 1, the most proximal E-box. We designed two PCR amplicons to assay for the presence of these two putative binding sites in chromatin immunoprecipitates. The experiments revealed that TWIST1 bound to E-box 1, but not to E-box 2 (Fig. 4e). Thus, TWIST1 specifically binds to the putative promoter of *mir-10b*, providing strong evidence that miR-10b can be directly regulated by this transcription factor.

Although miR-10b does not seem to be essential for cell motility in highly metastatic cancer cells (Fig. 1d), we asked nonetheless whether this miRNA is required for Twist1-induced migration and invasion in otherwise poorly motile cells. To this end, we introduced the antisense oligonucleotide for miR-10b into Twist1-overexpressing HMECs. In accord with a previous report[6], overexpression of Twist1 led to a strong increase in the motility and invasiveness of







f

	Lung metastasis	Peritoneal metastasis
Mock	0/10	0/10
miR-10b	8/10	3/10
P (χ^2 test)	0.01	0.1

Figure 3 | miR-10b induces distant metastasis. a, MECA-32- and Ki-67-stained sections of a primary mammary tumour formed by miR-10b-transduced SUM149 cells, at week 6 after orthotopic transplantation. Red arrows in panel A indicate tumour cells within a vessel, and black arrows in panel D indicate endothelial cells. Magnification: panels A and C, ×100; panels B and D: ×400. **b**, H&E- and AE1/AE3-stained sections of lungs isolated from mice that received orthotopic injection of miR-10b-transduced or mock-infected SUM149 cells, at week 9 after transplantation. Circles indicate clusters of metastatic cells. The arrow indicates normal bronchial epithelium. Inset, AE1/AE3 staining of a SUM149 primary tumour. Magnification, ×600. **c**, Numbers of lung micrometastases (micromets) per section in individual mice that received orthotopic injection of miR-10b-transduced or mock-infected SUM149 cells, at week 9 (left panel) and week 11 (right panel) after transplantation, respectively. **d**, Bright field, GFP imaging, and H&E staining of lungs isolated from mice that received orthotopic injection of miR-10b-transduced or mock-infected SUM159 cells, at week 11 after transplantation. Magnification, ×8 for bright field and GFP imaging; ×600 for H&E staining. **e**, GFP imaging and H&E staining of a macroscopic peritoneal metastasis in a mouse that received orthotopic injection of miR-10b-transduced SUM159 cells, at week 11 after transplantation. Magnification, ×8 for GFP imaging; ×400 for H&E staining. **f**, Incidence of lung metastasis and macroscopic peritoneal metastasis in mice that received orthotopic injection of miR-10b-transduced or mock-infected SUM159 cells.

these cells (Fig. 4f). Strikingly, miR-10b inhibition consistently led to a fivefold reduction in the motility and invasiveness of Twist1-overexpressing HMECs (Fig. 4f). Whereas Twist1 is capable, on its own, of inducing an EMT[6], miR-10b is not (data not shown). Instead, it seems to be essential to one element of the multi-component, Twist1-induced EMT program—increased cell motility and invasiveness.

HOXD10 is a direct and functional target of miR-10b

To understand the mechanisms by which miR-10b induces tumour invasion and metastasis, we used several computational methods to help identify miR-10b targets in humans. Among the approximately 100 targets predicted by both the TargetScan[32] and PicTar[33] search programs, two genes—homeobox D10 (*HOXD10*) and *RB1CC1* (also named *FIP200*)—were previously implicated in suppression of cell migration and/or invasion. *HOXD10* was of particular interest, because its expression has been found to be progressively lost in breast tumours showing increasing degrees of malignancy[34,35]. More importantly, restored expression of HOXD10 in MDA-MB-231 cells has been found to impair migration and invasion *in vitro* as well as tumour progression *in vivo*[35]. The *HOXD10*-encoded mRNA contains a 3′ UTR element that is partially complementary to miR-10b and carries the identical sequence in the human, mouse and rat mRNA orthologues (Fig. 5a).

Although miR-10b overexpression did not cause degradation of *HOXD10* mRNA (Fig. 5b), it did, however, reduce the activity of a luciferase reporter gene fused to the wild-type *HOXD10* 3′ UTR (48% reduction, $P = 0.003$, Fig. 5c), indicating that miR-10b targets *HOXD10* through translational inhibition. The action of miR-10b on *HOXD10* depends on the presence of a single miR-10b cognate binding site within the 3′ UTR, because the activity of a luciferase reporter that carries a mutant *HOXD10* 3′ UTR—with substitution of four nucleotides within the miR-10b binding site (Fig. 5a)—was not reduced by expression of miR-10b (Fig. 5c). In support of these results, we observed a clear reduction in the level of the endogenous HOXD10 protein in miR-10b-overexpressing cells (tenfold and threefold reduction in HMECs and SUM149 cells, respectively, Fig. 5d).

Others have demonstrated that HOXD10 represses expression of genes that are involved in cell migration and extracellular matrix remodelling, including RHOC, α3 integrin, matrix metalloproteinase-14, and urokinase-type plasminogen activator receptor[36]. Among these, RHOC has been identified as an especially important player

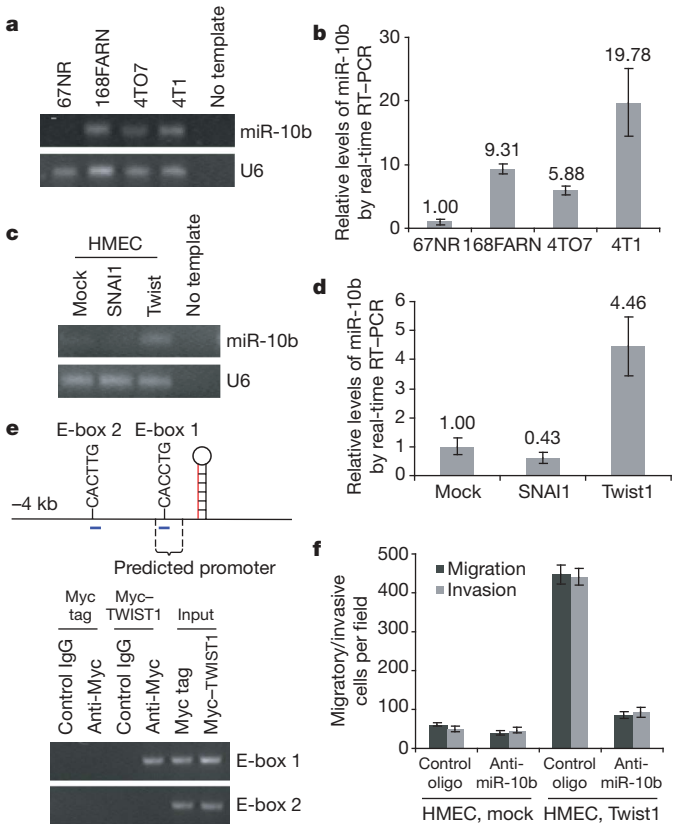


Figure 4 | miR-10b is regulated by Twist. **a**, **b**, RT–PCR (**a**) and real-time RT–PCR (**b**) of miR-10b in 67NR, 168FARN, 4TO7 and 4T1 cells. **c**, **d**, RT–PCR (**c**) and real-time RT–PCR (**d**) of miR-10b in HMECs transduced by SNAI1, Twist1 or the empty vector. **e**, Upper panel, human *mir-10b* genomic locus. The two short blue lines represent two PCR amplicons. Lower panel, ChIP assay in HEK293T cells transfected with a vector expressing Myc–TWIST1 or the Myc tag alone. PCR was performed with primers specific for human *mir-10b* E-box 1 and E-box 2, respectively. **f**, Transwell migration assay and Matrigel invasion assay of Twist1-transduced or mock-infected HMECs that were transfected with the inhibitor for miR-10b or the control oligonucleotide. A representative experiment is shown in triplicate along with s.e.m. in **b**, **d** and **f**.

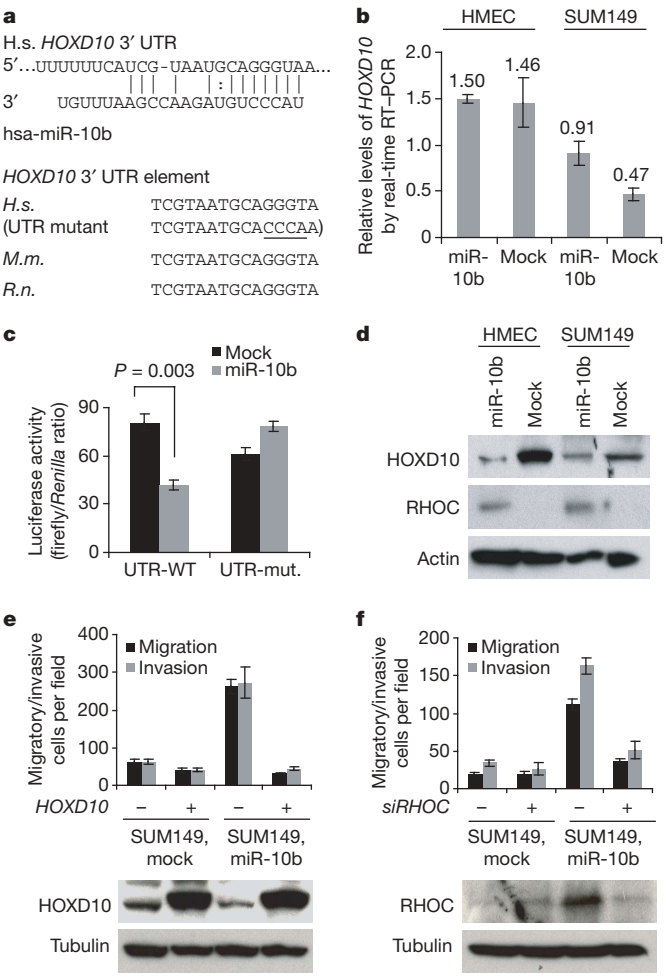


Figure 5 | miR-10b suppresses HOXD10, leading to induction of RHOC. **a**, Upper panel, predicted duplex formation between human *HOXD10* 3′ UTR and miR-10b. Lower panel, sequence of the miR-10b binding site within the *HOXD10* 3′ UTR of human (*H.s.*), and within the *HoxD10* 3′ UTR of mouse (*M.m.*), and rat (*R.n.*). **b**, Real-time RT–PCR of *HOXD10* in HMECs and SUM149 cells infected with the miR-10b-expressing or empty vector. Data were normalized to the level of *GAPDH* mRNA. **c**, Luciferase activity of wild-type (UTR-WT) or mutant (UTR-mut.) *HOXD10* 3′ UTR reporter gene in SUM149 cells infected with the miR-10b-expressing or empty vector. **d**, Immunoblotting of HOXD10 and RHOC in HMECs and SUM149 cells infected with the miR-10b-expressing or empty vector. **e**, **f**, Upper panel, transwell migration assay and Matrigel invasion assay of miR-10b-transduced or mock-infected SUM149 cells with transient transfection of *HOXD10* (**e**) or *RHOC* siRNA (**f**). Lower panel, immunoblotting of HOXD10 (**e**) or RHOC (**f**). A representative experiment is shown in triplicate along with s.e.m. in **b**, **c**, **e** and **f**.

in metastasis[37,38], and its expression correlates with metastatic spread of various types of carcinomas[39–41]. Indeed, we found that miR-10b-transduced cells exhibited robust expression of RHOC protein, whereas RHOC expression in the control cells was barely detectable (Fig. 5d).

We next ascertained whether reduction of HOXD10 levels might provide an explanation for the induction of cell motility and invasiveness observed following miR-10b overexpression. We overexpressed miR-10b in SUM149 cells together with a construct expressing HOXD10 constitutively; this construct encodes the entire HOXD10 coding sequence but lacks the 3′ UTR of HOXD10-encoding mRNA, yielding an mRNA that is resistant to miR-10b-mediated inhibition of translation. Strikingly, the resulting constitutive expression of HOXD10 completely abrogated miR-10b-induced cell motility and invasiveness (Fig. 5e), without affecting the proliferation or viability of these cells (data not shown), suggesting that this *HOX* gene is indeed a functionally important target of miR-10b. Furthermore, transfection of *RHOC* siRNA (small-interfering RNA), which caused a $>90\%$ reduction in the level of the RHOC protein (Fig. 5f), led to a strong but not complete suppression of miR-10b-induced cell migration (by 81%) and invasion (by 87%, Fig. 5f). Hence, RHOC seems to be a key downstream effector of miR-10b.

miR-10b expression is elevated in metastatic breast tumours

Paradoxically, a recent microarray study reported that miR-10b is among the miRNAs found to be downregulated in primary breast tumours (independent of their clinical aggressiveness) compared with normal breast tissue[22]. To address this apparent paradox and determine whether miR-10b expression correlates with clinical outcome in patients, we measured its levels in primary tumour samples from 23 breast cancer patients. When compared with normal breast tissue, miR-10b expression level was lower in all of the breast carcinomas from metastasis-free patients (5/5). In contrast, 50% of the metastasis-positive patients (9/18) had elevated miR-10b levels in their primary tumours ($P = 0.03$, Fig. 6a).

These results are in consonance with the expression pattern of miR-10b in cultured human mammary cells (Fig. 1c). In addition, we observed that ectopic miR-10b expression in non-tumourigenic, immortalized EpH4 mouse mammary epithelial cells[42] did not transform them to a tumourigenic state, whereas overexpression of Ras indeed enabled their tumourigenic growth (data not shown). Collectively, our findings indicate that miR-10b plays a part specifically in the metastatic process but not in primary tumour formation.

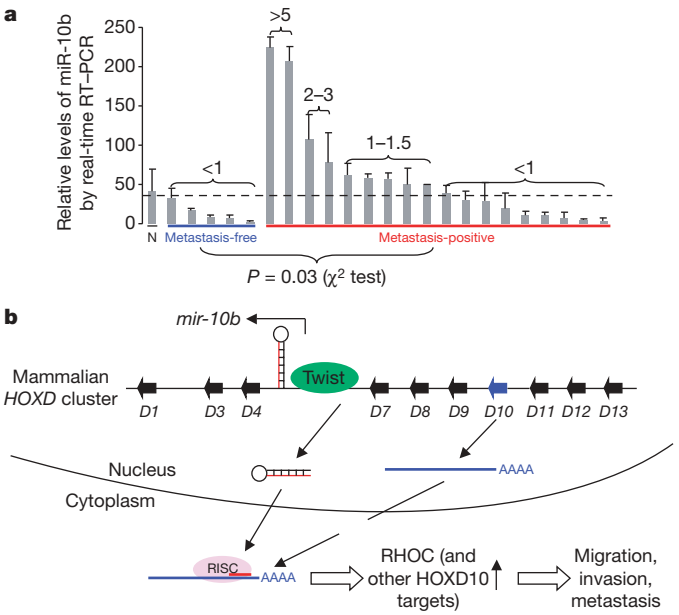


Figure 6 | miR-10b expression level is associated with the metastasis outcome in breast cancer patients. a, Real-time RT–PCR of miR-10b in 23 primary breast tumours from patients with indicated status of metastasis. N, normal breast tissue. Error bars indicate s.e.m. of triplicate experiments. The numbers above the columns indicate fold increase relative to normal tissue. **b**, Model for the regulation and function of miR-10b in cancer metastasis. RISC, RNA-induced silencing complex.

Discussion

The present work has led to the identification of a Twist-induced miRNA (miR-10b) that inhibits synthesis of the HOXD10 protein, permitting the expression of the pro-metastatic gene product, RHOC; this favours, in turn, cancer cell migration and invasion (Fig. 6b). Importantly, overexpression of this miRNA in otherwise non-metastatic breast cancer cells enables them to acquire invasive and metastatic behaviour. Conversely, silencing of miR-10b inhibits Twist-mediated cell migration and invasion. It remains to be seen whether miR-10b inhibition in highly malignant cells will reverse the metastatic phenotype *in vivo*. Such analysis is hampered at present by the lack of an experimental strategy for stably silencing miRNAs over extended periods of time.

Of particular interest, the *mir-10b* gene is located within the *HOXD* gene cluster (Fig. 6b). In fact, a number of *HOX* genes have been predicted to be targets of miRNA regulation, both in vertebrates[32] and in flies[43], and some of these associations have been validated experimentally: (1) miR-196, which has near perfect complementarity with the *HOXB8* mRNA 3′UTR region, downregulates *HOXB8* expression through cleavage of this mRNA[44]; and (2) miR-iab-4-5p downregulates ultrabithorax (UBX, a *Drosophila* HOX protein), leading to a homeotic phenotype in flies[45]. Accordingly, miR-10b represents a third 'HOX miRNA' that regulates expression of a *HOX* gene within the same *HOX* gene cluster that carries its encoding gene. This indicates that miR-10b is also likely to be involved in regulating certain steps of embryogenesis in which HOXD10 is known to participate. In addition, a close relative of miR-10b, miR-10a (the gene for which is located within the *HOXB* gene cluster), has been recently reported to target *HOXA1* (ref. 46), a gene that plays an oncogenic role in human mammary carcinoma cells[47], indicating that miR-10a might have an opposite rather than similar function in breast cancer.

In silico analyses predict that yet other mRNAs are direct targets of miR-10b-mediated silencing. Because the invasion–metastasis cascade involves a number of distinct steps[1], it is therefore possible that miR-10b regulates additional targets that are involved in several of these steps. A future challenge will be to identify the entire complement of miRNAs and their mRNA targets to elucidate more fully the contributions of these miRNAs to high-grade malignancy.

METHODS SUMMARY

Quantification of miRNAs was performed by real-time RT–PCR. Activity of miRNAs was measured by a luciferase reporter assay. Silencing of miRNAs was done by transfection of a chemically modified antisense oligonucleotide that was complementary to the endogenous miRNA being targeted. Ectopic expression of miR-10b was achieved by retroviral infection. Cell motility and invasiveness were gauged by transwell migration assay and Matrigel invasion assay, respectively. Chromatin immunoprecipitation was performed with a ChIP assay kit, followed by PCR with primers specific for *mir-10b*. Metastatic activity was assayed by mammary fat pad injection of cancer cells into NOD-SCID female mice. GFP imaging, histological analysis and cytokeratin immunohistochemistry were used to detect distant metastasis. Human materials were obtained from the Memorial Sloan-Kettering Cancer Center along with pathology reports and radiological reviews.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 June; accepted 16 August 2007.
Published online 26 September 2007.

1. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Rev. Cancer* **3**, 453–458 (2003).

2. Batlle, E. *et al.* The transcription factor snail is a repressor of *E-cadherin* gene expression in epithelial tumour cells. *Nature Cell Biol.* **2,** 84–89 (2000).
3. Cano, A. *et al.* The transcription factor snail controls epithelial–mesenchymal transitions by repressing *E-cadherin* expression. *Nature Cell Biol.* **2,** 76–83 (2000).
4. Comijn, J. *et al.* The two-handed E box binding zinc finger protein SIP1 downregulates *E-cadherin* and induces invasion. *Mol. Cell* **7,** 1267–1278 (2001).
5. Bolos, V. *et al.* The transcription factor Slug represses *E-cadherin* expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. *J. Cell Sci.* **116,** 499–511 (2003).
6. Yang, J. *et al.* Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117,** 927–939 (2004).
7. Hartwell, K. A. *et al.* The Spemann organizer gene, *Goosecoid*, promotes tumor metastasis. *Proc. Natl Acad. Sci. USA* **103,** 18969–18974 (2006).
8. Mani, S. A. *et al.* Mesenchyme Forkhead 1 (*FOXC2*) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proc. Natl Acad. Sci. USA* **104,** 10069–10074 (2007).
9. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nature Rev. Cancer* **6,** 857–866 (2006).
10. Esquela-Kerscher, A. & Slack, F. J. Oncomirs — microRNAs with a role in cancer. *Nature Rev. Cancer* **6,** 259–269 (2006).
11. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116,** 281–297 (2004).
12. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila. Cell* **113,** 25–36 (2003).
13. Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303,** 83–86 (2004).
14. Poy, M. N. *et al.* A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432,** 226–230 (2004).
15. Yi, R. *et al.* Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nature Genet.* **38,** 356–362 (2006).
16. Schratt, G. M. *et al.* A brain-specific microRNA regulates dendritic spine development. *Nature* **439,** 283–289 (2006).
17. Calin, G. A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl Acad. Sci. USA* **101,** 2999–3004 (2004).
18. He, L. *et al.* A microRNA polycistron as a potential human oncogene. *Nature* **435,** 828–833 (2005).
19. Johnson, S. M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120,** 635–647 (2005).
20. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834–838 (2005).
21. Roldo, C. *et al.* MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *J. Clin. Oncol.* **24,** 4677–4684 (2006).
22. Iorio, M. V. *et al.* MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* **65,** 7065–7070 (2005).
23. Meister, G., Landthaler, M., Dorsett, Y. & Tuschl, T. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA* **10,** 544–550 (2004).
24. Cheng, A. M., Byrom, M. W., Shelton, J. & Ford, L. P. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.* **33,** 1290–1297 (2005).
25. Elenbaas, B. *et al.* Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* **15,** 50–65 (2001).
26. Ethier, S. P., Mahacek, M. L., Gullick, W. J., Frank, T. S. & Weber, B. L. Differential isolation of normal luminal mammary epithelial cells and breast cancer cells from primary and metastatic sites using selective media. *Cancer Res.* **53,** 627–635 (1993).
27. Kuperwasser, C. *et al.* A mouse model of human breast cancer metastasis to human bone. *Cancer Res.* **65,** 6130–6138 (2005).
28. Thiery, J. P. Epithelial–mesenchymal transitions in tumour progression. *Nature Rev. Cancer* **2,** 442–454 (2002).
29. Cripps, R. M. *et al.* The myogenic regulatory gene *Mef2* is a direct target for transcriptional activation by Twist during *Drosophila* myogenesis. *Genes Dev.* **12,** 422–434 (1998).
30. Cheng, G. Z. *et al.* Twist transcriptionally up-regulates *AKT2* in breast cancer cells leading to increased migration, invasion, and resistance to paclitaxel. *Cancer Res.* **67,** 1979–1987 (2007).
31. Zhou, X., Ruan, J., Wang, G. & Zhang, W. Characterization and identification of microRNA core promoters in four model species. *PLoS Comput. Biol.* **3,** e37 (2007).
32. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115,** 787–798 (2003).
33. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nature Genet.* **37,** 495–500 (2005).
34. Makiyama, K. *et al.* Aberrant expression of *HOX* genes in human invasive breast carcinoma. *Oncol. Rep.* **13,** 673–679 (2005).
35. Carrio, M., Arderiu, G., Myers, C. & Boudreau, N. J. Homeobox D10 induces phenotypic reversion of breast tumor cells in a three-dimensional culture model. *Cancer Res.* **65,** 7177–7185 (2005).
36. Myers, C., Charboneau, A., Cheung, I., Hanks, D. & Boudreau, N. Sustained expression of homeobox D10 inhibits angiogenesis. *Am. J. Pathol.* **161,** 2099–2109 (2002).
37. Clark, E. A., Golub, T. R., Lander, E. S. & Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406,** 532–535 (2000).
38. Hakem, A. *et al.* RhoC is dispensable for embryogenesis and tumor initiation but essential for metastasis. *Genes Dev.* **19,** 1974–1979 (2005).
39. Kleer, C. G. *et al.* RhoC GTPase expression as a potential marker of lymph node metastasis in squamous cell carcinomas of the head and neck. *Clin. Cancer Res.* **12,** 4485–4490 (2006).
40. Kondo, T. *et al.* Expression of *RhoC* is associated with metastasis of gastric carcinomas. *Pathobiology* **71,** 19–25 (2004).
41. Wang, W. *et al.* Overexpression of the *RhoC* gene correlates with invasion and metastasis of hepatocellular carcinoma *Chinese J. Oncol.* (Zhonghua Zhong Liu Za Zhi.), **26,** 279–282 (2004).
42. Reichmann, E. *et al.* Activation of an inducible c-FosER fusion protein causes loss of epithelial polarity and triggers epithelial–fibroblastoid cell conversion. *Cell* **71,** 1103–1116 (1992).
43. Enright, A. J. *et al.* MicroRNA targets in *Drosophila. Genome Biol.* **5,** R1 (2003).
44. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* **304,** 594–596 (2004).
45. Ronshaugen, M., Biemar, F., Piel, J., Levine, M. & Lai, E. C. The *Drosophila* microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes Dev.* **19,** 2947–2952 (2005).
46. Garzon, R. *et al.* MicroRNA fingerprints during human megakaryocytopoiesis. *Proc. Natl Acad. Sci. USA* **103,** 5078–5083 (2006).
47. Zhang, X. *et al.* Human growth hormone-regulated *HOXA1* is a human mammary epithelial oncogene. *J. Biol. Chem.* **278,** 7580–7590 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Bartel, H. Lodish, P. Rao, B. Zhou, S. Mani, J. Yang, S. Ethier, C. Largman and L.-H. Wang for reagents and advice; F. Reinhardt for assistance with animal experiments; the Histology Core Laboratory at MIT and MSKCC for assistance with sectioning and immunohistochemistry; C. Mayr, C. Scheel, S. McAllister, I. Ben-Porath, Y. Sun and Y. Luo for critical reading of the manuscript; and members of the Weinberg Laboratory for useful discussions. L.M. is a Susan G. Komen Fellow of the Life Sciences Research Foundation. J.T.-F. is supported by the MSKCC Cancer Core Grant. R.A.W. is an American Cancer Society Research Professor and a Daniel K. Ludwig Cancer Research Professor. This research is supported by an NIH grant (R.A.W.) and the Ludwig Center for Molecular Oncology at MIT.

Author Contributions L.M. conceived the project. R.A.W. supervised research. L.M. designed and performed experiments. L.M. and J.T.-F. collected and analysed data. All authors contributed to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A.W. (weinberg@wi.mit.edu).

METHODS

Cell lines. The human mammary epithelial cells (HMECs, primary and immortalized) were described previously[25]. The MCF-10A, MCF-7, MDA-MB-231, and HEK293T cell lines were from ATCC and cultured under conditions provided by the manufacturer. Cell lines SUM149, SUM159 and SUM1315 were provided by S. Ethier and grown as described (http://www.asterand.com/Asterand/BIOREPOSITORY/AsterandSUMcelllineInstructions.pdf). 67NR, 168FARN, 4TO7 and 4T1 cell lines were cultured as previously described[6].

Constructs. The MDH1-PGK-GFP 2.0 vector was described previously[13]. The binding site for miR-9 or miR-10b, and the *HOXD10* 3′ UTR sequence were cloned into the pMIR-REPORT luciferase construct[24] (Ambion). The human *HOXD10* complementary DNA was a gift from C. Largman. Myc-tagged TWIST1 was provided by L.-H. Wang. The mutant construct of *HOXD10* 3′ UTR was generated using a QuikChange Site-Directed Mutagenesis Kit (Stratagene).

RNA isolation and miRNA detection. Total RNA from cultured cells, with efficient recovery of small RNAs, was isolated using the mirVana miRNA Isolation Kit (Ambion). Detection of the mature form of miRNAs was performed using the mirVana qRT–PCR miRNA Detection Kit and qRT–PCR Primer Sets, according to the manufacturer's instructions (Ambion). The U6 small nuclear RNA was used as an internal control.

miRNA gene cloning and ectopic expression. The human miRNA gene was PCR-amplified from normal genomic DNA and cloned into the MDH1-PGK-GFP 2.0 retroviral vector. The production of amphotropic viruses and infection of target cells were described previously[48]. The GFP^+ cells accounted for over 90% of infected cells, as determined by fluorescence-activated cell sorting (FACS) analysis.

Oligonucleotide transfection. The miRIDIAN microRNA inhibitors (Dharmacon) are single-stranded chemically enhanced oligonucleotides designed to inhibit the endogenous miRNAs. SMARTpool RHOC (Dharmacon) represents four pooled SMART-selected siRNA duplexes that target *RHOC* (the sense sequences are: GAAAGAAGCUGGUGAUCGUUU; GAACUAUAUUGCGGACAUUUU; GGACAUGGCGAACCGGAUCUU; CUACGUCCCUACUGUCUUUUU). Cells were transfected with 200 nM of the indicated oligonucleotide using the Oligofectamine reagent (Invitrogen). Forty-eight hours after transfection, cells were plated for migration and invasion assays, or harvested for the luciferase reporter assay.

***In vitro* migration and invasion assay.** For transwell migration assays, 2.5×10^4 to 5×10^4 cells were plated in the top chamber with the non-coated membrane (24-well insert; pore size, 8 μm; BD Biosciences). For invasion assays, 1.25×10^5 cells were plated in the top chamber with Matrigel-coated membrane (24-well insert; pore size, 8 μm; BD Biosciences). In both assays, cells were plated in medium without serum or growth factors, and medium supplemented with growth factors (for HMECs) or serum (for all other cells) was used as a chemoattractant in the lower chamber. The cells were incubated for 24 h and cells that did not migrate or invade through the pores were removed by a cotton swab. Cells on the lower surface of the membrane were stained with the Diff-Quick Staining Set (Dade) and counted.

Surgery, necropsy, histopathology and immunohistochemistry. Six-week-old female NOD-SCID mice (from Jackson Laboratory) were used for surgery. Mice were anaesthetized with 2,2,2-tribromoethanol. The skin was incised and tumour cells (1×10^6) in 25 μl growth medium were injected into the mammary fat pad. Each group consisted of 10 mice: for recipients of SUM149 cells, 3–4 mice per group were euthanized at each of the three time points (week 6, 9 and 11 post transplantation); for recipients of SUM159 cells, mice were euthanized when the tumours reached 2 cm in diameter, and thus the metastases by SUM159 cells were evaluated with mice carrying primary tumours of the same size, at the same time point (week 11–12 post transplantation). The mammary tumours were removed and weighed. The tumours, lungs, livers, spleens and macroscopic metastases were analysed under a dissecting microscope equipped with bright field and GFP fluorescence imaging. Tissue samples were fixed in 10% buffered formalin for 12 h, followed by a wash with PBS and transfer to 70% ethanol, and then embedded in paraffin, sectioned and stained with haematoxylin and eosin. The immunohistochemistry detection using anti-Ki-67 (1:50, Pharmingen) and anti-MECA-32 (1:50, from University of Iowa)-antibodies was performed on paraffin sections, using a BioGenex i6000 automated stainer in the Histology Core Laboratory at MIT. The immunohistochemistry detection using anti-cytokeratin antibodies AE1/AE3 (1:800, BioGenex) was performed on paraffin sections in the Histology Core Laboratory at MSKCC.

Luciferase reporter assay. Cells of 50% confluence in 24-well plates were transfected using Fugene6 (Roche). Firefly luciferase reporter gene construct (200 ng) and 1 ng of the pRL-SV40 *Renilla* luciferase construct (for normalization) were cotransfected per well. Cell extracts were prepared 24–48 h after transfection, and the luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega).

Chromatin immunoprecipitation (ChIP) assay. ChIP was performed with HEK293T cells transfected with a vector expressing Myc–TWIST1 or the Myc tag alone, using a ChIP assay kit (Upstate), according to the manufacturer's instructions, Protein–DNA complexes were precipitated with control IgG or anti-Myc antibody (Covance). PCR was performed with primers specific for human *mir-10b*.

Immunoblotting. Cells were harvested in RIPA lysis buffer (150 mM NaCl, 10 mM Tris, pH 7.5, 1% NP40, 1% deoxycholate, 0.1% SDS, protease inhibitor cocktail (Roche)). Proteins from total cell lysates were resolved by the NuPAGE 4–12% Bis-Tris gradient gel (Invitrogen), transferred to the PVDF membrane, blocked in 5% non-fat milk in PBS/Tween-20, and blotted with the antibodies for HOXD10 (1:200, Santa Cruz), RHOC (1:200, Santa Cruz), α-tubulin (1:10,000, Abcam), and β-actin (1:10,000, Abcam).

Patient study. Patient samples of primary breast carcinoma were consecutively ascertained at MSKCC between 1993 and 2005. Use of tissue samples was approved with an Institutional Review Board (IRB) Waiver and approved by the Human Tissue Utilization Committee (HTUC). All breast carcinoma samples were obtained at the time of mastectomy. Of the 18 patients diagnosed with metastases, all had lymph node metastases with documented pathology from total mastectomies with axillary node dissections. In addition, metastatic spread to the contralateral breast (3 occurences), chest wall (3), pleural fluid (1), and ovary (1), was documented by histological review of biopsies taken at the time of mastectomy and/or subsequent surgeries; diffuse osseous metastases and adnexal mass (1) and spinal metastases (4) were documented by radiological reviews including bone scans, CT scans, PET scans, and/or MRI scans. Total RNA, with efficient recovery of small RNAs, was isolated from four 20-μm sections from formalin-fixed, paraffin-embedded tissue blocks, using the RecoverAll Total Nucleic Acid Isolation Kit (Ambion). Quantification of miR-10b by real-time RT–PCR was performed as previously mentioned.

Statistical Analysis. Data are presented as mean ± s.e.m. Student's *t* test (two-tailed) was used to compare two groups ($P < 0.05$ was considered significant) unless otherwise indicated (χ^2 test).

48. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493–501 (2003).

ARTICLES

A histone H3 lysine 27 demethylase regulates animal posterior development

Fei Lan[1], Peter E. Bayliss[2]*, John L. Rinn[3]*, Johnathan R. Whetstine[1]*, Jordon K. Wang[3], Shuzhen Chen[1], Shigeki Iwase[1], Roman Alpatov[1], Irina Issaeva[4], Eli Canaani[4], Thomas M. Roberts[2], Howard Y. Chang[3] & Yang Shi[1]

The recent discovery of a large number of histone demethylases suggests a central role for these enzymes in regulating histone methylation dynamics. Histone H3K27 trimethylation (H3K27me3) has been linked to polycomb-group-protein-mediated suppression of *Hox* genes and animal body patterning, X-chromosome inactivation and possibly maintenance of embryonic stem cell (ESC) identity. An imbalance of H3K27 methylation owing to overexpression of the methylase EZH2 has been implicated in metastatic prostate and aggressive breast cancers. Here we show that the JmjC-domain-containing related proteins UTX and JMJD3 catalyse demethylation of H3K27me3/2. UTX is enriched around the transcription start sites of many *HOX* genes in primary human fibroblasts, in which *HOX* genes are differentially expressed, but is selectively excluded from the *HOX* loci in ESCs, in which *HOX* genes are largely silent. Consistently, RNA interference inhibition of *UTX* led to increased H3K27me3 levels at some *HOX* gene promoters. Importantly, morpholino oligonucleotide inhibition of a zebrafish *UTX* homologue resulted in mis-regulation of *hox* genes and a striking posterior developmental defect, which was partially rescued by wild-type, but not by catalytically inactive, human UTX. Taken together, these findings identify a small family of H3K27 demethylases with important, evolutionarily conserved roles in H3K27 methylation regulation and in animal anterior–posterior development.

Histone amino-terminal tails are subject to multiple post-translational modifications—including phosphorylation, ubiquitination, acetylation and methylation—and the combination of these modifications determines chromatin structure and transcriptional outcome[1]. Although histone methylation is linked to both repression and activation of transcription, histone H3K27 methylation has been considered to be critical for epigenetic repression[2]. H3K27 methylation is catalysed by the histone methyltranferase EZH2 (refs 3–6), which is the mammalian homologue of the *Drosophila* polycomb group protein Ez (enhancer of zeste). Both EZH2 and Ez have been shown to be essential for H3K27 trimethylation at the homeotic gene loci (*HOX*) and for silencing of the *HOX* genes[6,7], and therefore are important for body patterning along the anterior–posterior axis. EZH2 also has an important role in X-chromosome inactivation through regulation of H3K27 trimethylation[8,9]. H3K27 methylation has been shown to regulate transcription of tumour suppressor genes[10], and may also be important for stem-cell pluripotency regulation[11,12]. An imbalance of H3K27 methylation as a result of overexpression of EZH2 has been implicated in cancers[13,14]. Given the importance of H3K27 tri-methylation, an important issue to address was whether there are enzymes capable of reversing H3K27me3. Here we report the identification of two related JmjC-domain-containing proteins, UTX and JMJD3, as H3K27me3/2 specific demethylases, and provide evidence of a critical role for UTX in regulating H3K27 methylation at the *HOX* gene loci and in animal posterior development.

Discovery of a small family of H3K27me3 demethylases

Recent studies have identified JmjC-domain-containing proteins as histone demethylases that mediate the reversal of methylation at histone H3K4, H3K9 and H3K36 (refs 15–26). UTX, UTY and JMJD3 comprise a subfamily of JmjC-domain-containing proteins, which are evolutionarily conserved from *Caenorhabditis elegans* to human[15,16]. In addition to the JmjC domain, UTX and UTY, but not JMJD3, also contain tetratricopeptide repeats at their N-terminal regions, which are predicted to be protein interaction motifs (Supplementary Fig. 1a). All three proteins contain a treble-clef zinc finger[27] at their carboxy terminus. UTX resides on the X chromosome, escapes X-chromosome inactivation and is ubiquitously expressed[28]. UTY is a male-specific protein and may contribute to sex-specific tissue transplantation rejection response[29]. By and large, the biochemical and biological functions of these proteins were unclear.

Using recombinant UTX purified from Sf9 insect cells and a collection of methylated histone peptides as substrates, we found that UTX mediated demethylation of H3K27me3, but had no effect on the trimethylated H3K4, K9, K36 or H4K20 (Fig. 1a). UTX also mediated demethylation of H3K27me2 and H3K27me1 peptides (data not shown), indicating that UTX converts H3K27me3 to un-methylated products when assayed on peptide substrates *in vitro*. Consistently, western blot analysis also showed UTX-mediated demethylation of H3K27me3/2 on bulk histones (Fig. 1b). However, unlike the peptide substrates, a reduction of H3K27me1 levels was not detectable on the bulk histones (Fig. 1b, top panel).

In a similar manner to the full-length UTX, the UTX JmjC catalytic domain alone also mediated demethylation of H3K27me3, me2 and me1, when the methylated histone peptides were used as substrates (Supplementary Fig. 1b). When bulk histones were analysed, only H3K27me3 and me2, but not H3K27me1, levels were reduced (Fig. 1c). The JMJD3 catalytic domain displayed similar specificity,

[1]Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [2]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3]Program in Epithelial Biology, Department of Dermatology, Stanford University School of Medicine, 269 Campus Drive, Stanford, California 94305, USA. [4]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel.
*These authors contributed equally to this work.

demethylating H3K27me3, me2 and me1 of the histone peptides (Supplementary Fig. 1b), but only demethylating H3K27me3 and me2 on native histones and nucleosomal substrates (Fig. 1c, d). In contrast, the UTY JmjC domain or the full-length UTY purified from Sf9 cells was inactive under the same assay conditions (Fig. 1c, Supplementary Fig. 1b, and data not shown). Importantly, these enzymes showed no activity towards trimethylated H3K4, K9, K36 or H4K20 (Fig. 1a, c). The JMJD3-mediated demethylation reaction of the nucleosomal substrates is significantly less efficient than on bulk histones. Under the same assay condition, UTX had barely detectable activity (data not shown). This suggests that additional cofactors and/or posttranslational modifications may be required for efficient demethylation of nucleosomal substrates.

When overexpressed in 293T cells, JMJD3 significantly reduced the levels of H3K27me3 and me2 in approximately 78% and 56% of the transfected cells, respectively (Supplementary Fig. 2), but did not affect H3K36me3, H3K9me3 or H3K4me3 (Supplementary Fig. 2 and data not shown). Overexpression of JMJD3 did not reduce the H3K27me1 level, and, in fact, an increase in H3K27me1 was observed in 24% of the transfected cells (Supplementary Fig. 2). This accumulation of H3K27me1 is probably a result of the conversion of H3K27me3 and me2 to H3K27me1 due to overexpression of JMJD3. Overexpression of JMJD3 in HeLa cells, however, has less uniform effects on the global H3K27me3 level; that is, some transfected cells showed clear reduction of H3K27me3 but others showed less robust or marginal effects (data not shown). In contrast, overexpression of UTX in 293T cells had little effect, if any at all, on the global level of H3K27me3 (data not shown). These findings indicate that JMJD3 and UTX may have differential ability to impact global H3K27 trimethylation in a cell-type-specific manner and UTX action may be targeted to selective genomic loci in cells. Taken together, these findings suggest that UTX and JMJD3 are histone demethylases with specificities largely directed towards H3K27me3 and H3K27me2. Whether UTY is an active H3K27 demethylase remains to be determined by further experimentation.

UTX regulates H3K27 methylation at a *HOX* gene locus

We next asked whether UTX regulates H3K27 methylation at endogenous genes (Fig. 2). Because H3K27 trimethylation has been shown to be critical for the regulation of the *HOX* gene cluster, we investigated H3K27 methylation at the *HOX* gene locus and compared levels in HeLa cells in the presence and absence of short hairpin RNA (shRNA) plasmids that inhibited the expression of *UTX* or *JMJD3* (Fig. 2b). We found that RNA interference (RNAi) of the endogenous *UTX* resulted in H3K27me3 increases in some but not all *HOXD* genes. Specifically, in the promoters of *HOXD10* (primer no. 3), *HOXD11* (primer no. 7) and *HOXD12* (primers no. 10 and no. 11), H3K27me3 level was clearly elevated in the *UTX* knockdown cells (Fig. 2c). We also observed one instance in the *HOXD12* gene where H3K27me3 was elevated at the 3′ end of the gene in *UTX* knockdown cells (primer no. 9). In contrast, knockdown of *JMJD3* had very little effect on H3K27me3 levels within the *HOXD* cluster (Fig. 2c). These findings suggest that UTX is involved in regulating H3K27me3 levels at some but not all posterior *HOX* genes in HeLa cells. However, the increase in H3K27me3 levels was not accompanied by further repression of the corresponding *HOX* genes (data not shown). This may be due to the fact that these *HOX* genes are already highly silenced in HeLa cells.

Genome-wide analysis of UTX occupancy

To determine how general a role UTX plays in the regulation of the *HOX* gene loci, we carried out loci-specific and genome-wide analysis of UTX localization. We studied two primary human fibroblast cell types (foot and lung) that preserve distinct patterns of embryonic *HOX* expression[30], as well as mouse ESCs, in which *HOX* genes are silent[11]. The endogenous UTX in fibroblasts was isolated by chromatin immunoprecipitation (ChIP) followed by hybridization to

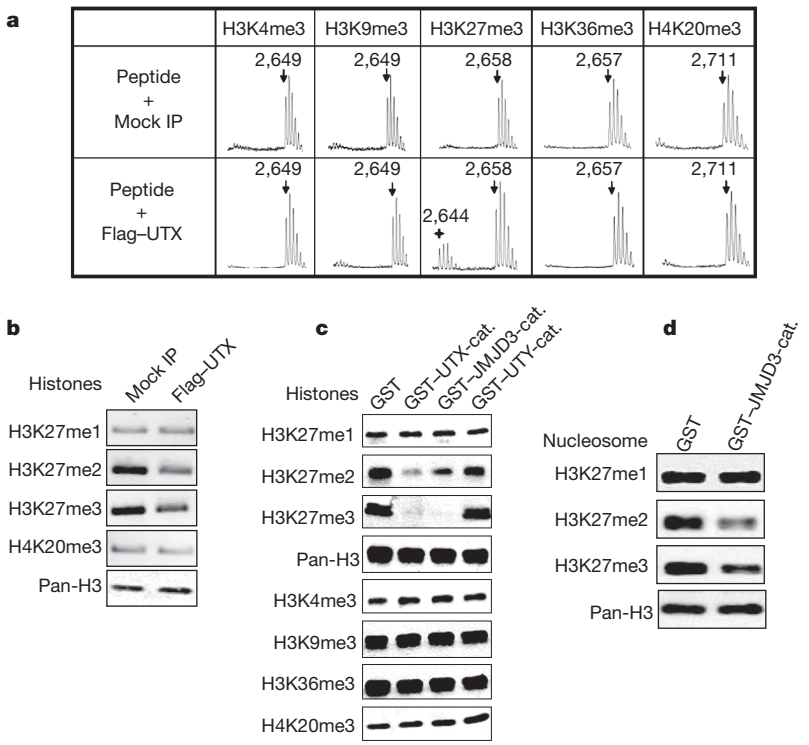


Figure 1 | Histone demethylation mediated by UTX. a, Purified UTX protein demethylates H3K27me3 but not tri-methylations on H3K4, 9, 36 and H4K20. Demethylated peptides are marked with a cross. Numbers refer to *m/z* values. **b**, UTX demethylates H3K27me3/2 on bulk histones. The demethylation events were detected by western blot using modification-specific antibodies. **c**, Demethylation of H3K27me3 and K27me2 on bulk histones by the catalytic domains of UTX and JMJD3. **d**, The JMJD3 catalytic domain demethylates H3K27me3/2 on mononucleosomal substrates. IP, immunoprecipitation; cat., catalytic domain; pan-H3 antibody detects total histone H3.

ultra-dense tiling microarrays (ChIP-chip) that interrogated all four human *HOX* loci at 5-base-pair resolution and 2 Mb of control regions, including portions of the X chromosome, chromosome 22, the β-globin locus, and many transcribed genes in fibroblasts[31]. Strikingly, UTX was selectively localized to the *HOX* loci of fibroblasts; over 90% of all UTX binding events were in the *HOX* loci (Fig. 3b, $P < 10^{-64}$, hypergeometric distribution). In contrast to the broad domains of H3K27me3, PRC2/SUZ12, H3K4me2 and RNA polymerase II occupancy over *HOX* genes in these cells, UTX was selectively targeted to narrow windows within 500 base pairs downstream of the transcriptional start site of *HOX* genes (Fig. 3a, c, and Supplementary Fig. 3). UTX binds the start of both transcriptionally active and silent *HOX* genes in a manner largely independent of anatomic origins of cells (Fig. 3a, d). Importantly, when UTX binding occurred within domains of H3K27me3, H3K27me3 was concomitantly diminished (Fig. 3c). This result supports the idea that UTX acts in cells as an H3K27me-specific demethylase to mark transcriptional start sites, and is consistent with the preferential effect of UTX depletion on H3K27me3 in gene-proximal promoters (Fig. 2).

Unlike fibroblasts with distinct positional identities, *Hox* genes in ESCs are largely occupied by H3K27me3 and transcriptionally silent[32]. We performed ChIP-chip analysis of UTX in mouse ESCs on tiling arrays that interrogated 4 kb surrounding ~48,000 annotated transcriptional start sites in the mouse genome. UTX was entirely excluded from the *Hox* loci (Fig. 3d), although UTX is expressed and seemed to bind selected sets of genes at other genomic locations (Supplementary Fig. 4). The lack of UTX occupancy indicates a potential deficiency in the mechanisms important for targeting of UTX to the *Hox* loci in mouse ESCs. To address the role of UTX binding to *HOX* genes, we depleted UTX in fibroblasts and examined the effects on *HOXA9*, a posterior *HOX* gene that bound UTX in a position-invariant manner. Depletion of UTX led to decreased expression of *HOXA9* in both lung and foot fibroblasts (Fig. 3e). In summary, these results suggest that endogenous UTX is actively and dynamically localized to the start of *HOX* genes in differentiation, potentially to reverse H3K27 methylation and to enable transcriptional activation.

Utx regulates zebrafish *hox* genes and posterior development

Given that UTX is located predominately at the *HOX* gene locus to regulate H3K27 methylation, we next asked whether UTX plays a role in animal body patterning, which is controlled by proper temporal and spatial expression of the developmentally regulated *Hox* genes. We turned to zebrafish, which possesses two UTX homologues (Supplementary Fig. 1). Inhibition of zebrafish Utx1 (NM_001080193) by two non-overlapping antisense morpholino oligonucleotides resulted in minor developmental delay by 24 h post fertilization (h.p.f.). By 48 h.p.f., posterior somites failed to retain the characteristic chevron shape (Fig. 4a). Inter-segmental distances are dramatically decreased in the caudal trunk, resulting in a much-shortened embryo and suggesting defects in posterior somitogenesis. Degeneration of the posterior notochord is also observed, from, approximately, somite 13 to 23, which results in a highly compact and shortened body at this affected region and a slight dorsal curvature of the tail (Fig. 4b). These phenotypes persist until 8 d.p.f. (data not shown). Anterior portions of the embryo seem mostly unaffected, and the head appears to develop normally, although it seems to be slightly larger than the control. Morpholino inhibition of fish *utx2* expression also affected posterior development, albeit to a much smaller extent (data not shown). Importantly, co-injection of 200 pg of messenger RNA encoding fish Utx1 or encoding human wild-type UTX1, but not that encoding catalytically inactive human UTX (Supplementary Fig. 5), resulted in a significant rescue of the development defects ($P = 5.1 \times 10^{-14}$, Fig. 4c, d, and data not shown). Consistently, there was no significant difference between the catalytically inactive mutant and the control in their ability to rescue the fish phenotype ($P = 0.26$). These findings further indicate that the phenotype was specifically due to Utx1 inhibition and suggest the demethylase activity is required for proper posterior patterning.

Given the finding of UTX occupancy at the *HOX* gene promoters and its role in H3K27me3 demethylation (Figs 2 and 3), and the fact that *HOX* genes are known to be important for body patterning during vertebrate development[33], we investigated in zebrafish whether *hox* genes are mis-regulated in *utx1* morphants. To avoid potential complications that may be caused by the size difference between the wild-type and *utx1*-morpholino-treated embryos at later time points, we carried out *hox* gene expression analysis at time points (12, 24 and 36 h post injection of the morpholinos) before the appearance of the posterior truncation phenotype. Importantly, we found that transcript levels of several 3′ posterior *hox* genes, such as *hoxa9b*, *hoxc8a*, *hoxc12b* and *hoxd12a*, are modestly but consistently reduced at 36 h d.p.f. (Fig. 4e). However, the expression of more anterior *hox* genes (such as *hoxc6a*, *hoxc6b* and *hoxa3a*)

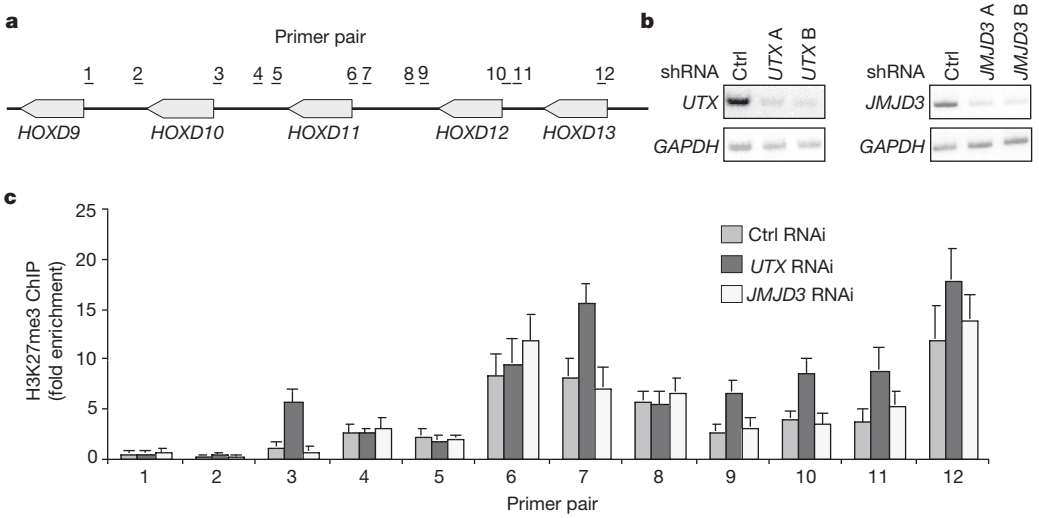


Figure 2 | UTX regulates H3K27me3 at the *HOXD* gene locus **a**, Schematic diagram of the mammalian *HOXD* gene cluster. The region covered by the primer sets used for ChIP are shown. The direction of transcription of the *HOX* genes shown in the diagram is from right to left. **b**, Specific shRNAs effectively knocked down the transcripts of *UTX* and *JMJD3* in HeLa cells (RT–PCR). **c**, ChIP analysis of H3K27me3 at the *HOXD* gene cluster. ChIP results with standard deviations using the H3K27me3 antibody and the primer sets of **a** are shown. ChIP results are expressed as the relative fold of enrichment at the *HOX* gene locus over a control region. Ctrl, control.

and the most posterior *hox* genes (*hoxa13b*, *hoxc13a* and *hoxc13b*) seems unaffected (Fig. 4e and data not shown), which corresponds well to the normal appearance of the anterior and most posterior tail regions of the mutant embryos (Fig. 4e and data not shown). Consistent with the PCR with reverse transcription (RT–PCR) results, RNA *in situ* analysis showed modestly reduced transcript levels of *hoxc8a* and a posterior shift of its expression domain in most of the *utx1*-morpholino-treated embryos (from somite 1–7 to somite 2–8; Fig. 4f). Similarly, we also observed reduced expression of *hoxc12b*, loss of expression of *hoxd12a* and *hoxa9b* at the pectoral fin bud, and a shift of the *hoxd9a* expression domain (Supplementary Fig. 6). Taken together, these findings indicate that Utx plays an important role in regulating *hox* gene expression and body patterning during vertebrate development through its demethylase activity, consistent with our ChIP-chip analyses, which identified UTX occupancy at *Hox* loci in differentiated cells (Fig. 3)

Recently, the ALR/MLL2 protein complex that mediates H3K4 trimethylation has been reported[34]. This complex contains both UTX and WDR5, the latter has been shown to be important for MLL-complex regulation of H3K4 methylation[35]. Interestingly, a posterior developmental defect and an improper expression pattern of *hoxc8*, similar to what we have observed for the *utx1*-morpholino-treated embryos, have also been noted in *wdr5*-morpholino-treated *Xenopus* embryos[35]. These findings indicate a potential collaboration of H3K4 methylation and demethylation of H3K27me in regulating *hox* gene expression. Consistently, we found co-localization of UTX with some of the other components of the ALR/MLL2 complex, including ALR/MLL2 and RBQ3/RBBP5, as well as H3K4me3, at

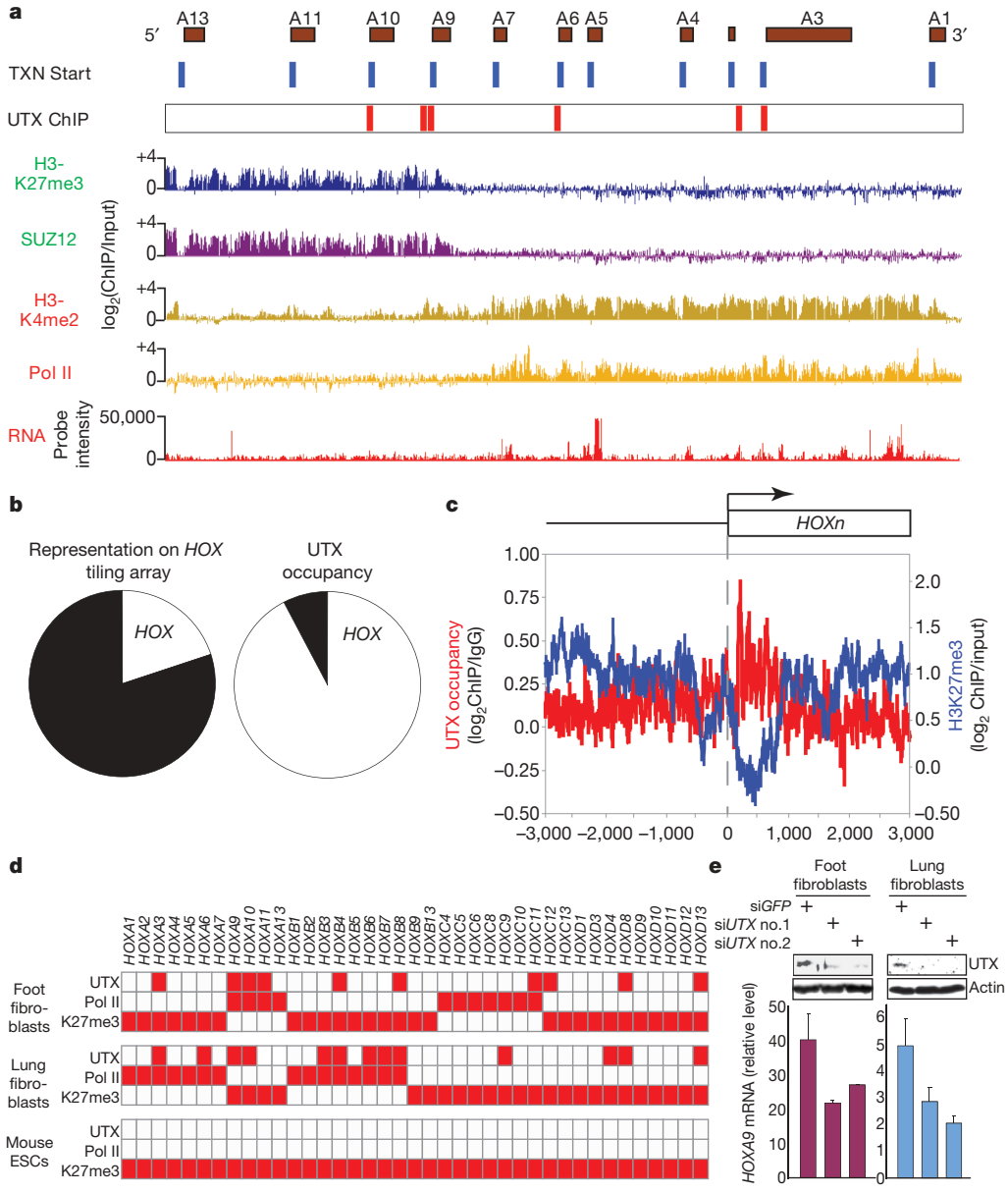


Figure 3 | UTX binds near *HOX* gene start sites to facilitate H3K27 demethylation and gene activation. **a**, *HOX* transcriptional start sites, the occupancy of UTX, H3K27me3, PRC2 (that is, SUZ12), H3K4me2 and RNA pol II, and RNA transcripts are shown over ~100 kb of the *HOXA* locus in lung fibroblasts. **b**, *HOX* genes represent ~20% of probes on the *HOX* tiling array but account for 92% of UTX binding events ($P < 10^{-64}$). **c**, Average UTX occupancy profiles for *HOX* genes bound by UTX aligned by *HOX* transcriptional start sites (left axis). Average H3K27me3 occupancy profiles of *HOX* genes bound by both UTX and H3K27me3 (right axis). The genes and ChIP profiles used are detailed in Supplementary Table 1. **d**, UTX is excluded from *HOX* loci of ESCs. Occupancy of UTX, PolII and H3K27me3 in ESCs and differentiated fibroblasts is shown as a matrix; red indicates occupancy. **e**, siRNA-mediated depletion of *UTX* decreases *HOXA9* transcription. Relative transcript levels (mean ± s.e.m.) by quantitative RT–PCR are shown. Upper panels, corresponding UTX and actin protein levels.

the *HOXA9*, *HOXA10* and *HOXD13* loci in lung fibroblasts, but not at other neighbouring regions (Supplementary Fig. 7). However, the collaboration of MLL2 and UTX may be cell-type-specific. In HeLa cells, the UTX–ALR/MLL2 complex does not occupy the *HOXA9* promoter, and consequently depletion of ALR/MLL2 does not affect H3K4 methylation or transcription of *HOXA9* (ref. 34). In addition, it is likely that H3K4 and H3K27 methylation can be toggled independently for certain genes. RNAi inhibition of *UTX* in human 293T or mouse ESCs had no effect on global H3K4 methylation or on the *in vitro* H3K4 methylation activity of ALR/MLL2 (Supplementary Fig. 8), and UTX occupancy did not correspond to broad H3K4me2 levels at the *HOX* loci (Supplementary Fig. 3c). In 293T cells, *HOXA9* is also regulated by the MLL1 complex, which is related to ALR/MLL2 but lacks UTX[36,37]. Taken together, these findings set the stage for future studies to delineate the relationship between MLL proteins and UTX, and the mechanism integrating the functions of the ALR/MLL2 complex in chromatin regulation.

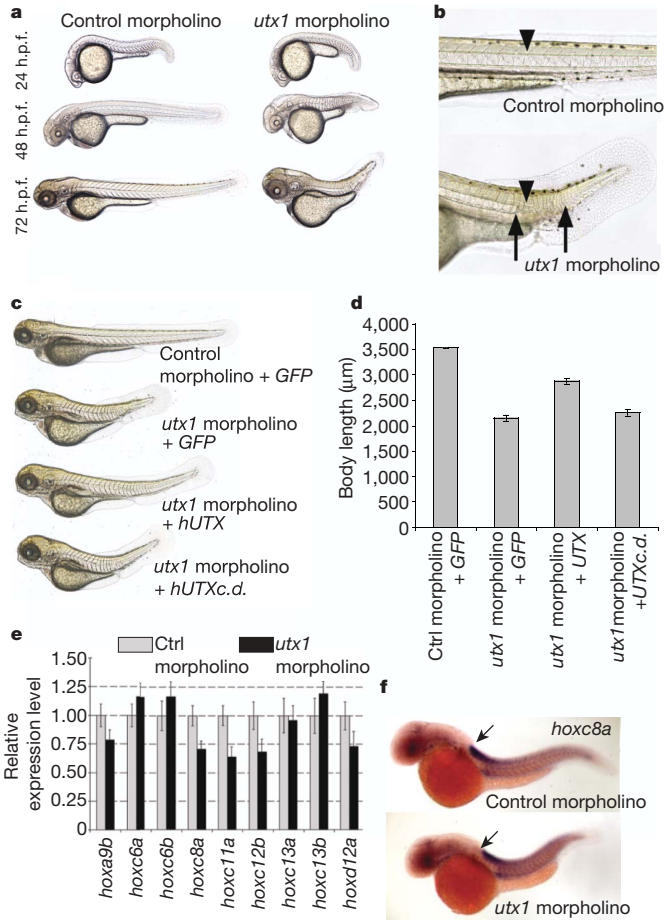


Figure 4 | Inhibition of zebrafish Utx1 results in improper development of the posterior trunk. a, Embryogenesis of the *utx* morphant was slightly delayed at 24 h.p.f. and a posterior development defect was observed after 48 h.p.f. **b**, Notochord degeneration (between arrows) was observed at 72 h.p.f. between, approximately, somites 13 to 23. In most knockdown embryos, the most distal portion of the notochord (between somites 24–30) remains fairly normal. **c**, **d**, Human *UTX* mRNA partially rescues the truncation phenotype of *utx1* morphants ($P = 5.1 \times 10^{-14}$). However, catalytically inactive *UTX* H1126A (*hUTXc.d.*) mRNA fails to rescue ($P = 0.26$). Error bars represent s.e.m. ($n = 50$). **e**, Quantitative RT–PCR analysis shows clear, though modest, reduction of several posterior *hox* genes in response to *utx1* morpholino. Error bars, s.e.m. **f**, Reduction of *hoxc8a* staining was observed in 6 out of 10 *utx1* morphants with a posterior shift of *hoxc8a* expression domains, which was observed in 4 out of the 10 embryos. Arrows point to the boundary between first and second somites. Purple, *hoxc8a* signal; orange, demarcates the somite boundaries (*dystrophin*).

Recent studies showed that a large set of developmentally important genes in ESCs are enriched for both H3K27me3 and H3K4me3 (bivalent domains), which represent the repressive and activating histone modifications, respectively[11]. The bivalent domains seem to segregate on differentiation, and the corresponding genes are either repressed or activated. Importantly, activation of some of these genes is correlated with a loss of H3K27me3 (ref. 32), indicating possible dynamic regulation through demethylation during differentiation. Our findings relating to the H3K27me demethylases, combined with the observation that JMJD3 expression is upregulated during ESC differentiation[32], supports the hypothesis that H3K27me demethylases have a role in the resolution of the 'bivalent domain' and in the regulation of transcription of these genes during ESC differentiation. Furthermore, our observation that UTX is excluded from the *Hox* gene loci identifies a possible mechanism that may help protect the bivalent domains at the *Hox* gene locus in ESCs. Given the fact that UTX is in the same protein complex with enzymes that mediate H3K4 trimethylation, we speculate that important mechanisms must be in place to facilitate differential regulation of the K4 and K27 methylation states in ESCs versus differentiated cells. These mechanisms may involve, but are not limited to, possible differential MLL/UTX complex composition and/or differential regulation of MLL and/or UTX enzymatic activities at the target loci at different stages of cell differentiation. Finally, a recent study has identified a link between retinoblastoma susceptibility gene *Rb* (also known as *RB1*), polycomb group proteins and H3K27 methylation in the transcriptional silencing of the $p16^{INK4\alpha}$ tumour suppressor[10]. *HOX* genes are also candidate oncogenes and tumour suppressor genes in several types of human cancer[38,39]. Thus, UTX and/or JMJD3 may play an antagonistic part to that of the polycomb group proteins in $p16^{INK4\alpha}$ and/or *HOX* regulation and therefore may function as a putative tumour suppressor.

In conclusion, we have identified a subfamily of H3K27 demethylases and among them UTX seems to have an important role in *HOX* gene regulation and animal body patterning. It has been proposed that H3K27me3 demethylases were either absent or were tightly regulated to protect H3K27me3, which is an important epigenetic modification[2]. The identification of H3K27me3 demethylases provides an additional layer of complexity for considering the regulation of this important epigenetic modification, and presents new opportunities to investigate how these demethylases are regulated to ensure both dynamic and stable epigenetic regulation.

METHODS SUMMARY

UTX, UTY and JMJD3 recombinant proteins were purified from bacteria with a glutathione *S*-transferase (GST) tag or from insect Sf9 cells with a Flag tag. We carried out *in vitro* demethylation assays using purified recombinant proteins with bulk histones, mono-nucleosomes or a panel of biotinylated histone peptides with specific methylations at indicated lysine residues. The demethylation reactions were detected by western blot with modification-specific antibodies or matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry. UTX and JMJD3 were knocked down by either short interfering (si)RNA or vector-based shRNA followed by antibiotics selection. Chromatin samples were prepared from indicated cells, and sonicated to the range of 200–600 bp. DNA samples from chromatin immunoprecipitation assays were analysed by radioactive duplex PCR or real-time PCR. Radioactive duplex PCR results were analysed by phosphor-image quantification and normalized to an internal control, which, in principle, is not regulated by H3K27me3 and H3K4me3. All PCR reactions were optimized to within the lineage range. UTX ChIP from human fibroblasts and mouse ESCs were analysed by a custom *HOX* tiling array, which interrogates all four *HOX* loci, two megabases of control regions and a mouse promoter tiling array set, respectively. Zebrafish Utx1 expression was inhibited by antisense morpholino injection at the one-cell stage. The body lengths of 50 embryos were measured at 96 h.p.f. by image analysis software (Spot Diagnostics). RNA samples from 20 control or *utx1* morphants (36 h.p.f.) were used to generate complementary DNA samples for quantitative RT–PCR analyses of *hox* genes. Whole-mount *in situ* hybridization was performed using digoxigenin and/or fluorescein-labelled antisense RNA probes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 August; accepted 23 August 2007.
Published online 12 September 2007.

1. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–45 (2000).
2. Trojer, P. & Reinberg, D. Histone lysine demethylases and their impact on epigenetics. *Cell* **125,** 213–217 (2006).
3. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039–1043 (2002).
4. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16,** 2893–2905 (2002).
5. Czermin, B. *et al. Drosophila* Enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111,** 185–196 (2002).
6. Muller, J. *et al.* Histone methyltransferase activity of a *Drosophila* Polycomb Group repressor complex. *Cell* **111,** 197–208 (2002).
7. Wang, J., Mager, J., Schnedier, E. & Magnuson, T. The mouse PcG gene *eed* is required for *Hox* gene repression and extraembryonic development. *Mamm. Genome* **13,** 493–503 (2002).
8. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300,** 131–135 (2003).
9. Heard, E. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Curr. Opin. Genet. Dev.* **15,** 482–489 (2005).
10. Kotake, Y. *et al.* pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing $p16^{INK4\alpha}$ tumor suppressor gene. *Genes Dev.* **21,** 49–54 (2007).
11. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
12. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125,** 301–313 (2006).
13. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419,** 624–629 (2002).
14. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl Acad. Sci. USA* **100,** 11606–11611 (2003).
15. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature Rev. Genet.* **7,** 715–727 (2006).
16. Shi, Y. & Whetstine, J. R. Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell* **25,** 1–14 (2007).
17. Yamane, K. *et al.* PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol. Cell* **25,** 801–812 (2007).
18. Iwase, S. *et al.* The X-linked mental retardation gene *SMCX/JARID1C* defines a family of histone H3 lysine 4 demethylases. *Cell* **128,** 1077–1088 (2007).
19. Tahiliani, M. *et al.* The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* **447,** 601–605 (2007).
20. Secombe, J., Li, L., Carlos, L. & Eisenman, R. N. The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes Dev.* **21,** 537–551 (2007).
21. Klose, R. J. *et al.* The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128,** 889–900 (2007).
22. Eissenberg, J. C. *et al.* The trithorax-group gene in *Drosophila* little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. *Nature Struct. Mol. Biol.* **14,** 344–346 (2007).
23. Christensen, J. *et al.* RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell* **128,** 1063–1076 (2007).
24. Lee, M. G., Norman, J., Shilatifard, A. & Shiekhattar, R. Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. *Cell* **128,** 877–887 (2007).
25. Liang, G., Klose, R. J., Gardner, K. E. & Zhang, Y. Yeast Jhd2p is a histone H3 Lys4 trimethyl demethylase. *Nature Struct. Mol. Biol.* **14,** 243–245 (2007).
26. Seward, D. J. *et al.* Demethylation of trimethylated histone H3 Lys4 *in vivo* by JARID1 JmjC proteins. *Nature Struct. Mol. Biol.* **14,** 240–242 (2007).
27. Ginalski, K., Rychlewski, L., Baker, D. & Grishin, N. V. Protein structure prediction for the male-specific region of the human Y chromosome. *Proc. Natl Acad. Sci. USA* **101,** 2305–2310 (2004).
28. Greenfield, A. *et al.* The *UTX* gene escapes X inactivation in mice and humans. *Hum. Mol. Genet.* **7,** 737–742 (1998).
29. Greenfield, A. *et al.* An H-YDb epitope is encoded by a novel mouse Y chromosome gene. *Nature Genet.* **14,** 474–478 (1996).
30. Rinn, J. L., Bondre, C., Gladstone, H. B., Brown, P. O. & Chang, H. Y. Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet.* **2,** e119 (2006).
31. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
32. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
33. Pearson, J. C., Lemons, D. & McGinnis, W. Modulating *Hox* gene functions during animal body patterning. *Nature Rev. Genet.* **6,** 893–904 (2005).
34. Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* **27,** 1889–1903 (2007).
35. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121,** 859–872 (2005).
36. Dou, Y. *et al.* Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature Struct. Mol. Biol.* **13,** 713–719 (2006).
37. Cho, Y. W. *et al.* PTIP Associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J. Biol. Chem.* **282,** 20395–20406 (2007).
38. Ayton, P. M. & Cleary, M. L. Transformation of myeloid progenitors by MLL oncoproteins is dependent on *Hoxa7* and *Hoxa9*. *Genes Dev.* **17,** 2298–2307 (2003).
39. Raman, V. *et al.* Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* **405,** 974–978 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements HeLa nucleosomes were a gift from X. Shi and O. Gozani. We thank G. Schnitzler for discussion. J.L.R. is a Fellow and H.Y.C. is the Kenneth G. and Elaine A. Langone Scholar of the Damon Runyon Cancer Research Foundation. J.R.W. is a recipient of the Ruth L. Kirschstein National Service Award. This work was supported by grants from the National Science Foundation (to J.K.W.), NIH, Scleroderma Research Foundation and Emerald Foundation (to H.Y.C), and from the NIH to Y.S. This work was also supported in part by a grant from the Novartis Biomedical Research Institute (to Y.S).

Author Information ChIP-chip data have GEO accession number GSE8855. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.S. (yshi@hms.harvard.edu).

METHODS

Peptides and antibodies. Biotinylated histone peptides were purchased from AnaSpec, and the sequence information is provided in Supplementary Table 2. Antibodies that recognize different histone modifications, H3K4me3 (05-745), H3K9me3 (07-442), H3k27me1 (07-448), H3K27me2 (07-452), H3K27me3 (07-449), H3K36me3 (07-549) and H4K20me3 (07-463), were purchased from Upstate Group. RBQ3 antibody (A300-109A) was purchased from Bethyl Laboratories, and UTX antibody was raised as previously described[34].

Plamid construction. Full-length *UTX*, *UTY* and *JMJD3* as well as the JmjC domains were PCR amplified from HeLa cells and a human male cDNA library using Phusion polymerase (Finnenzymes) and cloned into the pBlueBac4.5–Flag or pGEX4T-1 expression vectors, respectively.

Purification of recombinant proteins and demethylation reactions. The Flag-tagged UTX and UTY were expressed and purified from Sf9 cells using BAC-N-BLUE baculoviral expression system (Invitrogen) as previously described[18]. GST fused JmjC domains were purified from bacteria. Before elution, the affinity resins were washed twice with washing buffer without EDTA, which chelates Fe^{2+} and inhibits the demethylation reaction. For demethylation reactions, 2 μg of full-length proteins and 4 μg of JmjC domains were incubated with 1 μg of peptides or 4 μg of bulk histone in a 100 μl reaction for 4 h at 37 °C, as previously described[18].

MALDI–TOF mass spectrometry. One microlitre of the demethylation reaction mixture was desalted through a C18 ZipTip (Millipore). The ZipTip was activated, equilibrated, and loaded as previously described[18]. The bound material was then eluted with 10 mg ml^{-1} α-cyano-4-hydroxycinnamic acid MALDI matrix in 70% acetonitrile/0.1% TFA before being spotted and co-crystallized. The samples were analysed by a MALDI–TOF/TOF mass spectrometer.

Immunofluorescence microscopy. 293T and HeLa cells were transfected, fixed with 3% paraformaldehyde and immunostained as previously described[18]. The samples were mounted with Vectashield (Vector Laboratories) and analysed by fluorescence microscopy (Nikon E600) using a ×60 objective. Images were acquired and processed with Openlab 3.1.5 software.

RT–PCR. Total RNA was isolated with Trizol (Invitrogen) and reverse-transcribed with Superscript III (Invitrogen). For Fig. 4e, RNA samples were harvested from 15–20 control or *utx1* morphants. PCR reactions were conducted with equal amounts of cDNA and kept in the linear range. For Fig. 3e and Supplementary Fig. 8a, two independent siRNA duplexes against *UTX* (GGUCUUGGUUUGGUCUACUUCCAUU and UCAGUUAGCUUUGGUUGACUGUAAU) were purchased from Invitrogen (UTXHSS111233 and UTXHSS111234). A complete list of primers and conditions will be provided on request.

Chromatin immunoprecipitation. HeLa cells were transfected with shRNA vectors targeting *UTX* (GCCUCAGGAUGCCAUUAAA and GGGCAUGAACACAGUUCAA) and *JMJD3* (GGCGCUGACCAUUACCAAA and GCCACCAGGAGAAUAACAA) and selected by puromycin for 48 h. Cell pellets were resuspended in ChIP lysis buffer (50 mM HEPES/KOH, pH7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate and protease inhibitors) and chromatin was sonicated to the 200–600 bp range. Chromatin prepared from 10^7 cells was used in one ChIP experiment as previously described[34]. Eluted DNA samples were used in quantitative radioactive duplex PCR reactions (less than 30 cycles) and the results were obtained and analysed by the phosphor image system. The amplification was internally controlled by a reference region that is carefully selected. For Fig. 2c, the enrichment of H3K27me3 was calculated by [ChIP-*HOX*n/ChIP-Reference]/[INPUT-*HOX*n/INPUT-Reference]. For the reference regions, we used an intron region of the RNA Pol II gene or an intergenic region 10 kb upstream of the actin G gene where there is no H3K27me3 (Supplementary Table 3), which yielded similar results.

ChIP-chip. Primary human lung, foot and foreskin fibroblasts, custom human *HOX* tiling microarray, ChIP-chip analysis, and *HOX* loci-wide occupancy by SUZ12, H3K27me3, H3K4me2 and RNA pol II were as described[31]. ChIP of mouse ESCs (a gift from A. Wright and M. Scott) were hybridized to mouse promoter tiling array sets (Nimblegen Systems) that tile 3.25 kb upstream and 0.75 kb downstream of promoters, genome-wide. Affinity-purified UTX antiserum was used for ChIP as described above and compared with isotype-matched IgG in parallel ChIP-chip experiments. UTX occupancy was determined by a SignalMap peak-calling algorithm comparing binding events to simulated data on shuffled probe sets; we chose peaks that had estimated false-discovery rates of less than 0.01 and had signal intensity at least 1.5-fold over control chIP-chip experiments with IgG. ChIP-chip data are available for download from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE8855, and the Stanford Microarray Database (http://genome-www5.stanford.edu/).

Inhibition of the zebrafish UTX homologue and phenotypic analysis. Inhibition of zebrafish Utx1 was performed by injection of 2 nl of a stock concentration of 250 μM of antisense morpholinos (Gene-Tools, LLC) or 75 ng μl^{-1} of mRNA from human *UTX* constructs, or a control of enhanced green fluorescent protein (EGFP), into one-cell stage zebrafish embryos using a gas-driven microinjector (Medical Systems). The *utx1* morpholino sequence used in Fig. 4 was 5′-AGCTCCGAGCGTCCAAAAGCCACAA-3′ covering bases −55 to −31 in the 5′ UTR, and a similar phenotype was observed at a lower frequency with a second morpholino (5'-CCACCGACACTCGGCACGGCTTCAT-3′) covering ORF bases +1 to +25. The sequence of the control morpholino was 5′-CCTCTTACCTCAGTTACAATTTATA-3′. Whole-mount *in situ* hybridization was performed using digoxigenin and/or fluorescein-labelled antisense RNA probes (Roche) as previous previously described[40].

40. Chan, J. *et al.* Morphogenesis of prechordal plate and notochord requires intact Eph/ephrin B signaling. *Dev. Biol.* **234**, 470–482 (2001).

LETTERS

Association of the jets of Enceladus with the warmest regions on its south-polar fractures

Joseph N. Spitale[1] & Carolyn C. Porco[1]

Jets of material have been seen emanating from the south-polar terrain of Saturn's satellite Enceladus[1]. Observations have shown that this region is anomalously warm[2], with the hottest measured temperatures coinciding with the four 'tiger stripe' fractures, named Alexandria, Cairo, Baghdad and Damascus, that straddle the region[1,2]. Here we use Cassini images taken from a variety of viewing directions over two years to triangulate the source locations for the most prominent jets, and compare these with the infrared hotspot locations and the predictions from a recent model of tidally induced shear heating within the fractures[3]. We find that the jets emanate from the four tiger stripes, with the strongest sources on Baghdad and Damascus. All the jets from each fracture seem to lie in the same nearly vertical plane. There is a strong spatial coincidence between our geographical sources and the locations of increased temperature revealed by the infrared experiment. Comparison with the shear heating model shows broad agreement; the exception is the prediction that Baghdad is the least active lineament, whereas we find it to be the most active. We predict that several new hotspots remain to be discovered by future thermal observations.

Positional measurements of 'plumes' (that is, observed features, as opposed to the individual 'jets' of material that can compose a plume, depending on the viewing geometry) were taken from images[4] obtained from a variety of look directions with respect to the surface of Enceladus (Supplementary Table 2) using Narrow Angle Camera on board the Cassini spacecraft. A plume was measured in the two-dimensional image plane by selecting two points: one at the base of the visible plume, and another at higher altitude (Supplementary Figs 1–18). Each of the two points defines a geometrical ray extending from the camera, and the two rays define a plane in space coincident with the plume. The sources of all of the jets contributing to the measured plume must lie somewhere on the ground track formed by the intersection of the plume plane with the surface of the satellite (Fig. 1).

To isolate a particular jet viewed during various imaging sequences from the web of ground tracks derived from all of our images, we looked at the jet directions implied by the intersections of the various plume planes. Assuming that the source locations and the jet directions did not vary significantly throughout the data set, which spans two years, all of the inferred direction vectors as well as the ground-track intersections should be consistent for a given source. Because we found that nearly all of the ground tracks could be associated with some cluster of intersections possessing consistent directions, those assumptions seemed justified. A few of the remaining ground tracks occurred in clusters of two or three and fell in geographically interesting regions—near a tiger stripe or a Cassini Infrared Spectrometer (CIRS) hotspot—so we consider those to be sources as well, but with less confidence. See Supplementary Table 1 for specific details about the identification of each source.

Our solution for the source locations and jet directions consists of eight sources, each on one of the four tiger stripes. We can see a correspondence between our source locations and the hottest CIRS locales: of the six sources that occur in areas covered by the CIRS observations, three overlap CIRS hotspots and a further two are within less than one CIRS footprint of hotspots (Fig. 2). Because the latter two CIRS observations were not accompanied by simultaneous Imaging Science Subsystem imaging, the imperfect correspondence might be explained by errors in 'pointing', that is, locating the CIRS footprint on the surface. The source associated with the largest number of unambiguous sightings (Table 1), source I, was not observed at high resolution by CIRS; however, the CIRS low-resolution brightness temperature map[2] does suggest increased temperatures along the 30° W meridian in that vicinity.

We can also see a correlation between the CIRS temperature of each hotspot[2] and the number of unambiguous sightings of each jet throughout our data set (Fig. 2 and Table 1). This correlation may suggest a direct relationship between temperature and jet strength. If so, we would expect that the CIRS will measure higher temperatures at the location of source I than yet seen elsewhere on the surface, as well as somewhat increased temperatures at sources IV and VII.

The strongest sources (that is, those with the greatest number of sightings throughout our data set) are on Baghdad and Damascus. For each tiger stripe, although individual jets may emerge at oblique angles to the surface, all of the jets on that stripe seem to occupy nearly the same plane, which in all four cases is within a few degrees of being perpendicular to the surface. (Although only one source was identified on Alexandria, it is essentially vertical.) Each of these curtains of jets seems to be well collimated; a measurement of the material seen over the Baghdad stripe in set G suggests that it is confined to about $\pm 10^{\circ}$ of the mean plane, consistent with the scatter observed in the directions of the individual jets (Table 1).

If the largest discrepancies between our source locations and the CIRS hotspot locations—source II versus CIRS F and source V versus CIRS E—result from CIRS pointing errors, then future thermal observations with higher spatial resolution should place those hotspots closer to our source locations. However, tidal-stress calculations[5], which predict that jet activity should be greatest at times when rifts are in tension, show that the area encompassed by CIRS F would have been under compression during most of our observations. Therefore, if that model is correct, it is plausible that the CIRS pointing is accurate and the region instead periodically produces a jet that we were unlucky enough to miss. Indeed, the counts in Table 1 show that jets were more likely to be observed at times when the corresponding regions were predicted to be in tension than in compression. That comparison is clouded by the poor statistics and by the fact that individual jets are often obscured by other jets, depending on the viewing geometry. However, one notable supporting example is provided by the observation that source I, our most obvious source, is

[1]Cassini Imaging Central Laboratory for Operations (CICLOPS), Space Science Institute, 4750 Walnut Street, Boulder, Colorado 80301, USA.

not apparent in set K, despite near-optimum viewing geometry. At the time of that observation, the region around source I was in compression. A more rigorous investigation of the time variability of the jetting and overall plume brightness is planned for a future paper.

Another explanation for the above discrepancies, in addition to the fact that we see no material emanating from CIRS G, might be a change in the jetting activity at those locations during the past millennium in accordance with the 1,000-year cooling timescale arising from thermal simulations[6]; that is, the sub-surface source of the heating may have turned off, but the residual thermal heat has not yet dissipated.

If the planes containing the jets that we have determined for each tiger stripe reflect the orientations of the corresponding fissures, then those fissures are essentially vertical. That orientation may be consistent with their formation as tensile cracks under the current stress regime, because the current orientation of the tiger stripes corresponds to nearly maximum normal tidal stress[3]. Alternatively, jetting has been observed elsewhere in the Solar System with no apparent confining vent[7], so the collimation may instead arise from other physical processes. However, this non-confined jetting was vertical and, although our plume planes are vertical, not all of the individual jets contained in those planes emerge vertically.

Localized heating resulting from tidally driven strike–slip motion along the tiger stripes has been proposed as a source of the anomalous heat radiating from the tiger stripes, which in turn powers the jets[3]. That calculation predicts maximum heating on Alexandria around 150° W longitude (very close to source IV), on Cairo around 60° W longitude close to source V (CIRS E) as well as at 120° W near source VIII (CIRS B), on Damascus near sources II and III, and on the lower latitude branch of Damascus near CIRS F. Minimal heating is predicted on Baghdad, although a small region near source VI (CIRS C) does show a significant signal. There are a few additional small regions where that model predicts increased heating and where we see no jets. Thus, although that calculation successfully predicts most of our observed jets, it fails to explain the fact that we observe Baghdad to be the most active tiger stripe on the basis of the sighting counts in Table 1 and supported by increased CIRS brightness temperatures in that vicinity. It does corroborate the location of CIRS F, lending additional support to the idea that it is the site of a jet that was in the dormant part of its cycle at the times of our observations, instead of a mis-registered CIRS footprint. It also corroborates source II. Therefore, if the pointing for CIRS F is accurate, then the

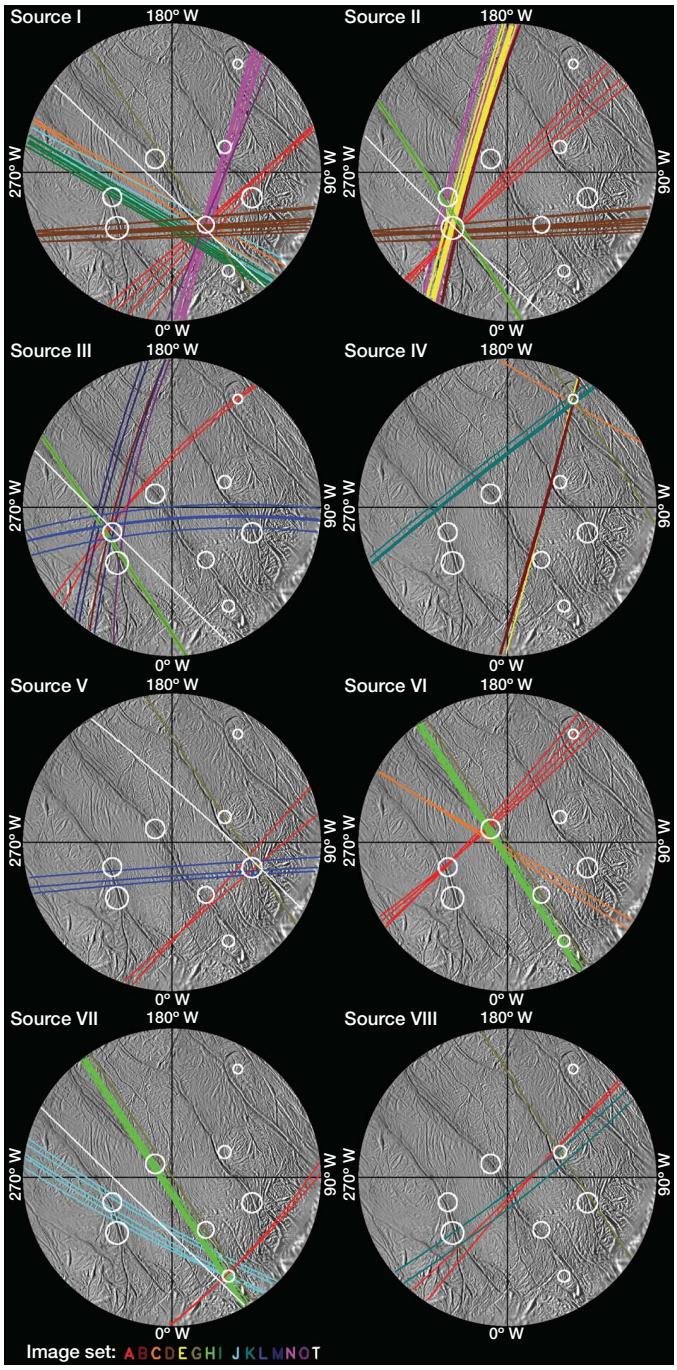


Figure 1 | Stereographic projections with ground tracks for each source solution. Each image set is denoted by a unique colour. Circles encompass the standard deviation of the cluster of intersections comprising each source. Some ground tracks are shown for more than one source because the unique source cannot be distinguished (for example, sets G and H, viewing along the tiger stripes). Source numbers are shown on the larger map in Fig. 2. Where necessary to bring out detail, image sets were averaged before measuring, and hence show only one ground track (see Supplementary Figs 1–18).

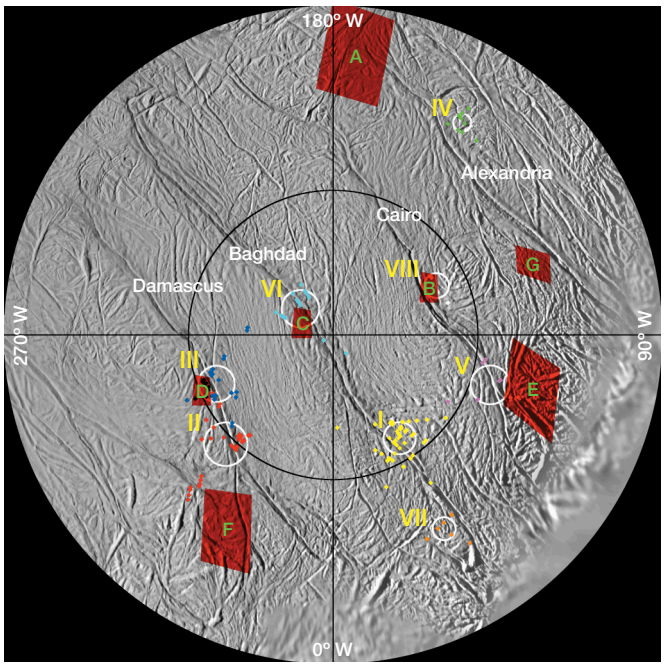


Figure 2 | Stereographic projection of the south-polar terrain showing our sources and CIRS hotspots. Our source locations are labelled with yellow roman numerals; CIRS hotspots are labelled with green capital letters. The intersections comprising each source are shown as coloured diamonds. Circles represent the standard deviations of those sets of intersections. Intersections of planes separated by less than 20° were excluded because the precise location of those intersections is overly sensitive to small variations in the plane orientations.

Table 1 | Jet source locations, directions, sighting counts and CIRS hotspot associations

Source	Tiger stripe	Lat. (°)	Lon. (° W)	σ (°)	Azimuth (°)	Zenith (°)	σ (°)	#T	#C	#CT	#TC	Sightings	CIRS (ref. 2)	Temperature (K) (ref. 2)
I	Baghdad	−81.5	32.8	1.6	(228.7)	9.7	9	3/5	1/2	2/3	1/1	7	–	–
II	Damascus	−79.4	315.5	2.1	(93.8)	6.8	9	3/5	0/0	0/0	0/0	5	(F)	157 ± 24
III	Damascus	−81.3	292.8	1.8	83.2	30.2	10	3/6	0/2	2/2	0/0	6	D	145 ± 14
IV	Alexandria	−72.9	148.7	0.7	(110.0)	3.7	15	0/0	1/3	1/1	0/0	4	–	114 ± 22
V	Cairo	−78.6	72.3	1.7	(229.8)	6.0	5	0/3	3/6	0/0	0/0	3	E	133 ± 28
VI	Baghdad	−87.1	231.4	1.8	(187.6)	10.2	2	1/4	1/2	0/0	0/0	2	C	135 ± 9
VII	Baghdad	−74.6	29.8	1.1	352.5	20.8	5	2/6	0/3	0/0	0/0	3	–	–
VIII	Cairo	−82.1	115.5	1.0	(127.7)	6.8	12	1/3	0/4	0/2	0/0	2	B	117 ± 16
													G	127 ± 28

Latitudes (Lat.) are planetographic and longitudes (Lon.) increase toward the west. Azimuths are measured clockwise from local north (that is, eastward to an observer standing on the surface facing north); values in parentheses are poorly determined owing to the small zenith angle (angle measured from the vertical). The uncertainties σ corresponds to the standard deviation in the spherical coordinates corresponding to the set of vectors comprising each quantity (source location or jet direction). Columns labelled with # record the number of times a source was unambiguously detected in a given stress state[5]—T, tension; C, compression; CT, transitioning from compression to tension; and TC, transitioning from tension to compression—compared with the number of times it might have been seen in that state in our data set. This count excluded sightings from sets G and H because that geometry (looking along the tiger stripes) did not allow us to determine which particular jets were active on a given tiger stripe at those times. The set D observations were also excluded because they cannot be linked uniquely to either source I or II. We counted sightings in multiple images within a single image set only once. Configurations for which it was difficult to determine the prevailing stress state at that location at that time were not counted. Totals for all sources are given in the bottom row. CIRS hotspot associations in parentheses are questionable. '–' indicates that there is no associated hotspot.

region containing source II was never observed by CIRS and we would expect an additional hotspot to be found there.

Received 17 May; accepted 30 August 2007.

1. Porco, C. C. *et al.* Cassini observes the active south pole of Enceladus. *Science* **311,** 1393–1401 (2006).
2. Spencer, J. R. *et al.* Cassini encounters Enceladus: background and the discovery of a south polar hot spot. *Science* **311,** 1401–1405 (2006).
3. Nimmo, F. *et al.* Shear heating as the origin of the plumes and heat flux on Enceladus. *Nature* **447,** 289–291 (2007).
4. Porco, C. C. *et al.* Cassini imaging science: instrument characteristics and anticipated scientific investigations at Saturn. *Space Sci. Rev.* **115,** 363–497 (2004).
5. Hurford, T. A. *et al.* Eruptions arising from tidally controlled periodic openings of rifts on Enceladus. *Nature* **447,** 292–294 (2007).
6. Spencer, J. R. *et al.* The south polar hot spot on Enceladus. *Lunar Planet. Sci. Conf.* **XXXVII,** 2252 (2006).
7. Milazzo, M. P. *et al.* Observations and initial modeling of Lava-SO2 interactions at Prometheus, Io. *J. Geophys. Res.* **106 E12,** 33121–33128 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Paul and Pauline Helfenstein for the planning of the Enceladus plume observations and the staff members at the CICLOPS for technical support. D. Wilson at CICLOPS provided assistance with image processing. We thank H. J. Melosh, A. P. Ingersoll and P. Helfenstein for their comments. This work was funded by the Cassini project.

Author Contributions Both authors devised the approach and contributed equally to the discussion and conclusions. J.N.S. executed the work, produced the solutions and wrote most of the text. C.C.P. instigated the work and revised the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.N.S. (spitale@lpl.arizona.edu).

LETTERS

Phase-resolved measurements of stimulated emission in a laser

Josef Kröll[1], Juraj Darmo[1], Sukhdeep S. Dhillon[2,3], Xavier Marcadet[4], Michel Calligaro[4], Carlo Sirtori[2] & Karl Unterrainer[1]

Lasers are usually described by their output frequency and intensity. However, laser operation is an inherently nonlinear process. Knowledge about the dynamic behaviour of lasers is thus of great importance for detailed understanding of laser operation and for improvement in performance for applications. Of particular interest is the time domain within the coherence time of the optical transition. This time is determined by the oscillation period of the laser radiation and thus is very short. Rigorous quantum mechanical models[1,2] predict interesting effects like quantum beats, lasing without inversion, and photon echo processes. As these models are based on quantum coherence and interference, knowledge of the phase within the optical cycle is of particular interest. Laser radiation has so far been measured using intensity detectors, which are sensitive to the square of the electric field. Therefore information about the sign and phase of the laser radiation is lost. Here we use an electro-optic detection scheme to measure the amplitude and phase of stimulated radiation, and correlate this radiation directly with an input probing pulse. We have applied this technique to semiconductor quantum cascade lasers, which are coherent sources operating at frequencies between the optical ($>$100 THz) and electronic ($<$0.5 THz) ranges[3]. In addition to the phase information, we can also determine the spectral gain, the bias dependence of this gain, and obtain an insight into the evolution of the laser field.

Since the realization of the first laser, researchers have investigated stimulated emission and the gain in an amplifying medium using coherent or incoherent radiation[1]. Several well known phenomena have been observed in these very basic experiments—spectral narrowing, power broadening, gain saturation, and the spectral gain curve[4–7]. With the advent of ultrashort laser pulses, time-resolved spectroscopy with a time resolution better than 10 fs has become possible[8]. At the same time, femtosecond-laser-driven generation of few-cycle phase-locked terahertz (THz) pulses has paved the way towards coherent detection of transient THz electric fields using electro-optic sampling[9]. This latest generation of phase-resolved femtosecond spectroscopy has recently enabled the observation of the birth of the plasmon-phonon mode[10] and of excitons in semiconductors[11], as well as the observation of gain associated with an optically pumped excitonic system in Cu_2O (ref. 12).

The phase-resolved measurement of stimulated emission in lasers has however still been hindered by the rather fast electric field oscillations in conventional lasers (about 2.5 fs at 800 nm). The situation has changed dramatically with the realization of THz quantum cascade lasers (THz-QCL) emitting in a wide frequency range from 4.4 THz down to 1.6 THz (refs 13, 14). Accordingly, the time resolution required to resolve the electric field oscillations of the THz-QCL emission has been brought down to several tens of femtoseconds, which is readily accessible by modern femtosecond spectroscopy.

With the set-up described in the Methods section we have direct access to the amplitude and phase of the THz pulses transmitted through the laser. Knowledge of the phase is essential when studying optical processes in an inverted system. From the quantum mechanical description of stimulated emission, it is known that incoming photons generate a superposition between ground state and first excited state. This superposition constitutes an oscillating dipole that emits electromagnetic waves. The difference between amplification and attenuation is solely determined by the phase of this superposition. When the incoming and the dipole fields are in-phase, constructive interference and amplification will be observed. Destructive interference and thus attenuation of the transmitted pulse will be observed when these fields are out-of-phase. In the present detection scheme, we resolve the phase of the superposition directly from the relative phase between probe and emitted electric fields.

The pulse transmitted through the unbiased THz-QCL (Fig. 1a) resembles the input pulse followed by small oscillations, which point to a weak absorption in the THz-QCL waveguide. When the laser is biased, the shape of the differential transmission of the THz pulse changes significantly (Fig. 1b). The time domain signal is now a result of changed absorption and gain in the biased THz-QCL. The transient THz field is dominated by a very pronounced oscillation lasting several picoseconds. These oscillations are due to stimulated emission and thus amplification of the incident pulse. It can be seen that there is a phase jump between the oscillating part and the initial pulse, as indicated by an arrow in Fig. 1b. In Fig. 1c we compare these data to a finite-difference time-domain simulation of the pulse propagation in the two-level system[15,16] described in Methods and Supplementary Information. We find excellent agreement with respect to the envelope and phase of the oscillations if we assume a population inverted system with a gaussian distribution of the resonant frequencies centred at 2.9 THz, a coherence time T_2 of 7 ps, and a non-radiative decay time of 1.5 ps.

From the discrete Fourier transformation of the modulation signal, we obtain the absorption spectrum and phase (Fig. 2a and b, respectively). The absorption is dominated by a negative peak at about 2.9 THz, representing the spectral gain of the THz-QCL. The frequency dependence of the phase—displaying the negative part on the low-frequency side of the resonance—is inverted compared to the usual behaviour for resonant absorption. This behaviour is the unambiguous evidence for gain and thus population inversion. With knowledge of the phase, we can distinguish between amplification and loss. Thus we can attribute the broadband positive feature at frequencies between 1.0 and 2.2 THz to increased loss in

[1]Photonics Institute, Vienna University of Technology, Gusshausstrasse 25-29, A-1040 Vienna, Austria. [2]Matériaux et Phénomènes Quantiques, Université Paris 7, 75251 Paris Cedex 05, France. [3]Ecole Normale Superieure, 75231 Paris Cedex 05, France. [4]Alcatel-Thales III-V Lab, Route Départementale 128, 91767 Palaiseau Cedex, France.

the THz-QCL caused by differences in the carrier distribution in the unbiased and biased active region. This broadband absorption causes the initial pulse of the THz transient in Fig. 1b. In Fig. 2c, the spectral gain is shown for different bias current densities. Below threshold, we observe an almost gaussian gain profile with a bandwidth (full-width at half-maximum) of about 130 GHz. Above threshold (at 165 A cm^{-2}), the gain curve is changed and we observe a dip in the gain spectrum. We explain this by gain clamping at the frequency of the Fabry–Perot cavity mode. The position of this hole coincides exactly with the Fabry–Perot mode at 2.87 THz of the lasing QCL, which was measured separately using a Fourier transform infrared spectrometer. The observation of the dip in the spectral gain is possible because the THz-QCL is working in a single mode condition (for a discussion of the interaction between a laser field and the spatial distribution of gain, see ref. 17). Note that we are observing the gain spectrum while the laser is above threshold. This measurement is unique, and illustrates the enormous potential of our technique.

The potential of electro-optic sampling is further demonstrated by looking at the pulses that are reflected at the facets of the Fabry–Perot resonator of the QCL. These reflected pulses are temporally separated by the cavity round-trip time. We perform these measurements below threshold at a current density of 110 A cm^{-2} to avoid the above-mentioned effects of spatial gain modulation in the QCL lasing at the Fabry–Perot modes. This would cause a more complex evolution of the multiple reflected pulses. Staying below threshold provides a homogeneous gain and allows a straightforward explanation of the results. Figure 3 displays pulses after a few passes through the cavity. The amplitude of the pulse after three passes (Fig. 3b) is comparable to that of the direct transmission (Fig. 3a), but the pulse duration is significantly extended. The initial feature corresponding to the broadband response of the incident pulse is completely gone. After five passes (Fig. 3c), the amplitude of the oscillation appears to be lower, the duration of the pulse is again longer and the total energy in the pulse is slightly less than that of the three-pass pulse. The observed pulse duration tendency points to the spectral narrowing of the pulses. The energy carried by the individual pulses decreases because the total propagation losses prevail over gain when the laser is driven below threshold.

Finally, we focus on the bias current dependence of the single-pass gain and correlate it with the light output and the current–voltage characteristics (Fig. 4). This comparison reveals details about the internal processes in the active region of the QCL. Without applied bias or with low bias, the THz-QCL exhibits no resonant absorption at the lasing frequency (~2.9 THz) usually expected for an unpumped gain medium. However, without the correctly applied bias the energy levels in the QCL are not properly aligned[18], thus the designed optical transition cannot occur and no resonant absorption can be observed. A measurable gain appears only when most of the cascades in the active region are properly aligned, which is indicated by a change in the differential resistance at a bias of ~ 2.1 V (Fig. 4c). After the onset, the gain rises proportionally with bias current density owing to an increasing population inversion. The lasing threshold is reached at a bias current density of about 113 A cm^{-2}, where we obtain a single-pass gain of ~19 cm^{-1}. This value compares to calculated waveguide losses of 7–11 cm^{-1} (refs 19–21) and to a value of 12 ± 1 cm^{-1} obtained from measurements of unbiased THz-QCL waveguides at different lengths. We assume

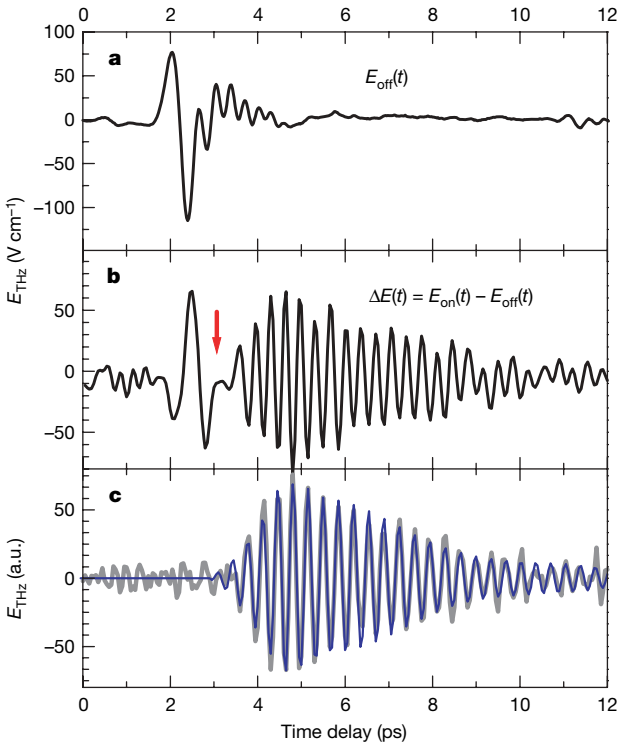


Figure 1 | Terahertz time-domain spectroscopy of a THz-QCL. **a**, The electric field waveform, $E_{off}(t)$, of a THz pulse transmitted through an unbiased THz-QCL. **b**, Change in the transmitted pulse waveform, $\Delta E(t)$, induced by biasing the THz-QCL. The signal is dominated by strong oscillations. A phase jump occurs between the initial pulse and the oscillations indicated by an arrow. **c**, Comparison of a simulation (thin blue line) to $\Delta E(t)$ spectrally filtered by a band-pass filter 2.0–3.5 THz (grey line). The leading part of the terahertz pulse disappears because it carries only broad-band low frequency components below 2 THz. The bias current density was 110 A cm^{-2}.

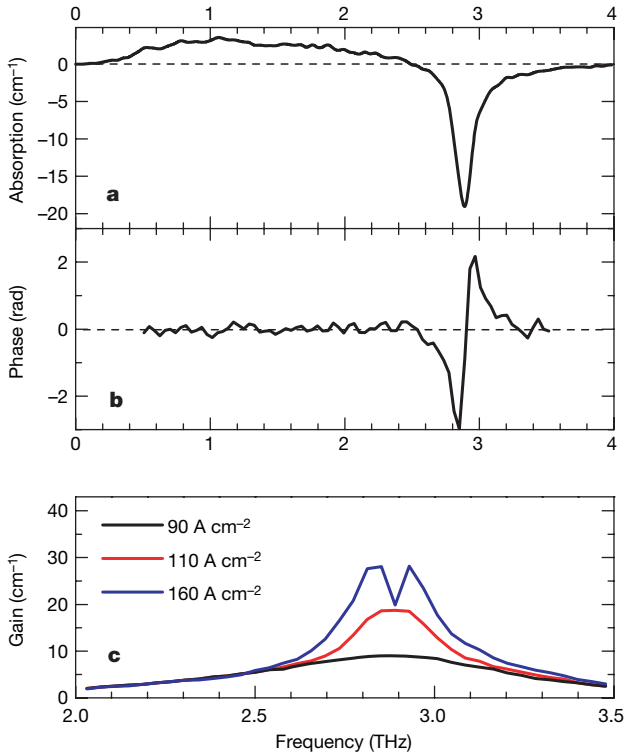


Figure 2 | Fourier transform of the transmission signal modulated by the THz-QCL. **a**, The absorption shows a negative part (gain) and positive part (reduced transmission). **b**, The phase change retrieved from the THz electric field shows an inverted dispersive behaviour, as expected from an inverted system. **c**, Spectral gain of the THz-QCL driven at 90, 110 and 160 A cm^{-2}. For the highest bias current density, the device is lasing at a single Fabry–Perot mode. We observe a dip in the spectral gain corresponding to the frequency of the Fabry–Perot mode. For this frequency, the gain is clamped at the threshold value.

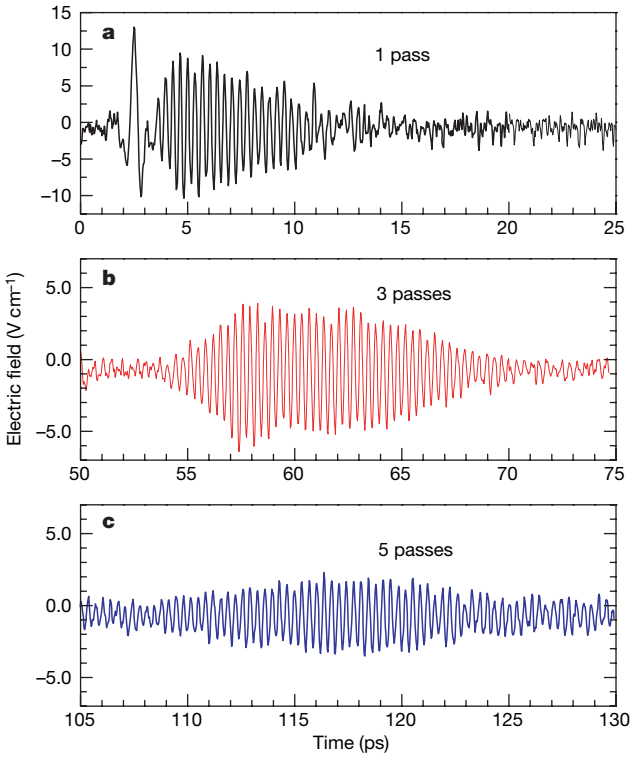


Figure 3 | Build-up of the electric field oscillations phase-locked to the external source. **a**, Waveform of the directly transmitted (single pass) THz pulse. **b**, The pulse waveform after three passes through the cavity. The pulse amplitude is comparable to that of the direct transmission, but the duration is significantly extended and the initial part of the pulse has vanished. **c**, After five passes, the amplitude of the oscillation appears to be lower, but the pulse length is now much higher—the total energy in the pulse is almost the same as for the three-pass pulse. The THz-QCL is operated slightly below threshold at a current density of 110 A cm^{-2}. The pulses are temporally separated by the cavity round-trip time.

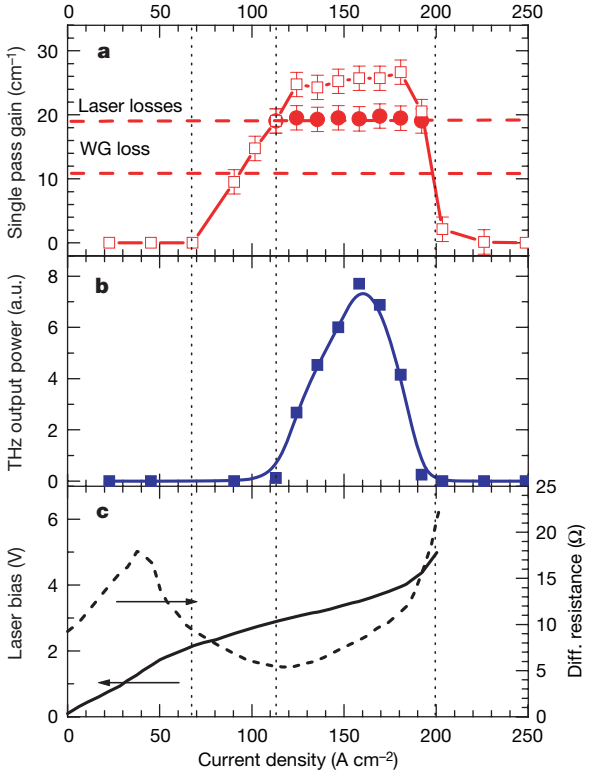


Figure 4 | Bias current characteristics of the THz QCL. **a**, Single-pass gain obtained from the electro-optic measurement at the spectral hole (2.87 THz, filled symbols) corresponding to the saturated gain, and evaluated at the maximum of the spectral gain (2.95 THz, open symbols) representing the unsaturated gain. WG, waveguide. Error bars indicate standard deviation. **b**, Average output power of the QCL measured by a photo-acoustic detector, showing the threshold and the dynamic range. **c**, Current–voltage curve (solid line) and differential resistance (dashed line) of the QCL. The QCL is operated in pulsed mode at a repetition rate of 17 kHz and a duty cycle of 15%.

mirror losses as high as 7 cm^{-1} because of a silicon hemispherical coupling lens attached to one waveguide facet.

The bias current dependence of the single-pass gain changes significantly when the lasing threshold is reached, which is detected by a sudden increase of the total THz output intensity recorded by the intensity detector (a Golay cell; Fig. 4b). The single-pass gain of the THz pulse, however, does not increase linearly any more in this current range, but it becomes almost 'clamped' at the threshold value. The standing-wave pattern of the laser mode leads to a local gain clamping at the antinodes, while the gain remains unclamped at the nodes. The THz pulse passing through the laser cavity probes both saturated and unsaturated gain areas, resulting in the apparent hole in the gain spectrum (Fig. 2c). We can therefore determine the unsaturated and clamped gain values (Fig. 4a).

In contrast to diode lasers, the gain of the THz-QCL is maintained for a limited dynamical range (Fig. 4a). This is explained by a gradual decrease of the injection into the upper lasing level due to field-induced shifts of the injector levels[18]. Associated with this effect is a clearly observable steep increase of the QCL's differential resistance above bias current densities >180 A cm^{-2} (Fig. 4c).

With phase-resolved measurements of the gain in laser media, a whole class of new experiments can be performed and fundamental physical parameters accessed. By noting the sign of the phase, it is possible to unambiguously determine gain or loss. Using both electro-optic field detection and intensity measurements provides the basis for making simultaneous observations of amplification and self-sustained lasing. Access to these separate processes allows the study of waveguide or other optical losses, as well as the examination of nonlinear effects like saturation, gain clamping, and spatial hole burning. The technique we have presented is not limited to QCLs, but is applicable to any gain medium. With the recent development of phase-locked extreme-ultraviolet pulse generation, the electric field on the attosecond timescale can be resolved[22], allowing the technique we have presented here to be used in the infrared and visible spectral range.

METHODS SUMMARY

We use THz time-domain spectroscopy[23] to access the coherent temporal properties of the THz-QCL. Quasi-single-cycle THz pulses, generated by exciting a photoconductive switch[24] with 80-fs-long, infrared laser pulses from a Ti:sapphire laser, are transmitted through the THz-QCL waveguide. Coherent electro-optic detection[9] is used to sense the instantaneous electric field vector of the transmitted THz pulses. Complementary to the electro-optic detection we use also a photo-acoustic detector (a Golay cell). This detector registers the total THz intensity emitted from the laser, while the electro-optic detector is sensitive only to the THz radiation which is phase-locked with respect to the incident THz pulse. The measurements are performed on an AlGaAs/GaAs THz-QCL with a bound-to-continuum optical transition designed for 2.9 THz (refs 19, 20). Details of the experimental technique and the design of the THz-QCL are given in Methods and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 March; accepted 20 August 2007.

1. Scully, M. O. & Lamb, W. E. Quantum theory of an optical maser. I. General theory. *Phys. Rev.* **159**, 208–226 (1967).
2. Haken, H. Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. *Rev. Mod. Phys.* **47**, 67–121 (1975).
3. Faist, J. *et al.* Quantum cascade laser. *Science* **264**, 553–556 (1994).

4. Bennet, W. R. Jr. Hole burning effect in a He-Ne optical maser. *Phys. Rev.* **126,** 580–593 (1962).
5. Rigrod, W. W. Gain saturation and output power of optical masers. *J. Appl. Phys.* **34,** 2602–2609 (1963).
6. Osgood, R., Eppers, W. & Nichols, E. An investigation of the high-power CO laser. *IEEE J. Quant. Electron.* **QE-6,** 145–154 (1970).
7. Crowe, W. & Ahearn, W. F. Semiconductor laser amplifier. *IEEE J. Quant. Electron.* **QE-2,** 283–289 (1966).
8. Diels, J.-C. & Rudolph, W. *Ultrashort Laser Pulse Phenomena: Fundamentals, Technique, and Applications on a Femtosecond Time Scale* (Academic, San Diego, 1996).
9. Wu, Q. & Zhang, X.-C. Ultrafast electro-optic field sensor. *Appl. Phys. Lett.* **68,** 1604–1606 (1996).
10. Huber, R. *et al.* How many-particle interactions develop after ultrafast excitation of an electron–hole plasma. *Nature* **414,** 286–289 (2001).
11. Huber, R., Kaindl, R. A., Schmid, B. A. & Chemla, D. S. Broadband terahertz study of excitonic resonances in the high-density regime in $GaAs/Al_xGa_{1-x}As$ quantum wells. *Phys. Rev. B* **72,** 161314 (2005).
12. Huber, R. *et al.* Stimulated terahertz emission from intraexcitonic transitions in Cu_2O. *Phys. Rev. Lett.* **96,** 017402 (2006).
13. Koehler, R. *et al.* Terahertz semiconductor-heterostructure laser. *Nature* **417,** 156–159 (2002).
14. Walther, C., Scalari, G., Faist, J., Beere, H. & Ritchie, D. Low frequency terahertz quantum cascade laser operating from 1.6 to 1.8 THz. *Appl. Phys. Lett.* **89,** 231121 (2006).
15. Taflove, A. *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, Boston, 1995).
16. Ziolkowski, R. W., Arnold, J. M. & Gogny, D. M. Ultrafast pulse interactions with two-level atom. *Phys. Rev. A* **52,** 3082–3094 (1995).
17. Kröll, J. *et al.* Longitudinal spatial hole burning in terahertz quantum cascade lasers. *Appl. Phys. Lett.* (in the press).
18. Faist, J., Capasso, F., Sirtori, C., Sivco, D. L. & Cho, A. Y. in *Intersubband Transitions in Quantum Wells: Physics and Device Applications II* (eds Liu, H. C. & Capasso, F.) 1–83 (Academic, London, 2000).
19. Alton, J. *et al.* Buried waveguides in terahertz quantum cascade lasers based on two-dimensional plasmon modes. *Appl. Phys. Lett.* **86,** 071109 (2005).
20. Barbieri, S. *et al.* 2.9 THz quantum cascade lasers operating up to 70 K in continuous wave. *Appl. Phys. Lett.* **85,** 1674–1676 (2004).
21. Kohen, S., Williams, B. & Hu, Q. Electromagnetic modeling of terahertz quantum cascade laser waveguides and resonators. *J. Appl. Phys.* **97,** 053106 (2005).
22. Baltuska, A. *et al.* Attosecond control of electronic processes by intensive light fields. *Nature* **421,** 611–615 (2003).
23. Nuss, M. C. & Orenstein, J. in *Millimeter and Submillimeter Wave Spectroscopy of Solids* (ed. Gruener, G.) Ch. 2 (Springer, Berlin, 1998).
24. Zhang, X.-C., Hu, B. B., Darrow, J. T. & Auston, D. H. Generation of femtosecond electromagnetic pulses from semiconductor surfaces. *Appl. Phys. Lett.* **56,** 1011–1013 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to D. P. Kelly for comments on the presentation of the manuscript. We acknowledge financial support from the European Commission under the Integrated Project TeraNova funded by the IST directorate, and from the Austrian Science Fond (FWF) under the project Advanced Light Sources (ADLIS).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.U. (karl.unterrainer@tuwien.ac.at).

METHODS

Spectroscopy and lasers. We use standard THz time-domain spectroscopy[8] to access the optical properties of the THz-QCL. In our specific case (Supplementary Fig. 1a), broadband THz electromagnetic pulses are generated by a gallium arsenide (GaAs) based photoconductive emitter[24] excited by 80 fs pulses at 800 nm from a Ti:sapphire mode-locked laser (pulse repetition rate is 80 MHz). The THz pulse is injected into the THz-QCL waveguide through one of the laser's facets, where it propagates along the laser cavity axis, and is re-emitted at the second facet into free space. To ensure sufficient coupling efficiency, we use a solid immersion lens approach (Supplementary Fig. 1b) to focus the THz radiation into the THz-QCL waveguide. A silicon hemisphere lens is attached directly to the laser facet. In addition, a square metallic aperture ($200 \times 200\ \mu m^2$) is placed between lens and waveguide to mask the THz transparent substrate of the laser chip.

The instantaneous electric field vector of the THz pulse is sensed using an electro-optic sampling technique based on the Pockels effect[9]. We use a 300-μm-thick gallium-phosphide (GaP) nonlinear optical crystal as electro-optic active medium. With this combination of a GaAs emitter and a GaP detector, a typical frequency spectrum between 0.5 and 3.5 THz is covered with a signal dynamic range >20 dB. Since the electro-optic detection is a coherent detection technique, only the THz radiation, which has a constant phase shift with respect to the sampling femtosecond pulse, is recorded. This phase-locked measurement is different in principle from standard transmission measurements[2–5] performed on lasers until now. Several major features emphasize the novelty of the method: (1) amplitude and phase of the amplified pulse's electric field are measured, in contrast to the standard time-averaged intensity measurements made using a probing beam; (2) broadband probing in the spectral range from 0.5 to 3.5 THz; (3) inherently high time resolution (<80 fs), which allows monitoring of the processes in the laser cavity on a subpicosecond timescale.

Response of system to THz pulse. The response of the THz medium to the injected THz pulse is obtained by examining the difference between the THz electric field transmitted through the THz-QCL, with and without bias. In order to suppress parasitic effects of scattered light and of electrical interferences in the set-up, the THz emitter and the laser are modulated at two different frequencies, ω_1 and ω_2, respectively, and the measured signal is then detected at the difference frequency $\omega_1 - \omega_2$. The spectral amplitude and phase information are obtained by standard Fourier transformation of the time domain data. The typical frequency resolution of the set-up is 30 GHz. To complement the electro-optic detection data, we use also a photo-acoustic detector (Golay cell) to record the total THz field intensity.

All the measurements are performed on AlGaAs/GaAs THz-QCL with a bound-to-continuum optical transition[19,20]. The active region of the THz-QCL consists of 90 cascades, each 125 nm long. The band structure of one period (with applied electric field) is shown in Supplementary Fig. 1c. The laser devices are processed into surface plasmon ridges 110 μm wide and 2 mm long. The laser threshold current density is $113\ A\,cm^{-2}$, the device emits single mode at 2.87 THz and operates up to a sink temperature of 70 K. All parameters refer to pulsed operation at a repetition rate of 17 kHz and a duty cycle of 15%.

The interaction of a quasi-single-cycle pulse with the optical transition in a QCL can be modelled via the interaction of an electromagnetic wave with a two-level quantum system. This approximation of a THz-QCL is adequate if we consider the lowest energy level of the injector region as the upper level of the two-level system (optical transition) and the highest energy level of the next injector region as the lower level of the two-level system (Supplementary Fig. 1c). In addition, we assume that both energy levels interact with other levels in the heterostructure via tunnelling and scattering at timescales much longer than the optical transition rate. Thus the two-level quantum system representing the optical transition can be—within a certain timescale—considered as an isolated system. We use Maxwell–Bloch equations[16] to describe the interaction of the THz pulse with this system. The results of the finite-difference time domain simulation[20] are summarized in Supplementary Fig. 2. The electric field transient forces a superposition of the wavefunctions of the ground and excited states of the two-level system and a temporal oscillating dipole is formed. The electric field emitted by this dipole is out of phase with the excitation field when the ground state population dominates, $n_a > n_b$ (Supplementary Fig. 2a). If a population inversion $n_a < n_b$ is set, the emitted electric field and the driving field are in phase (Supplementary Fig. 2b). The amplitude of the emitted field is proportional to the population difference between the ground and excited state ($n_a - n_b$) and the emission vanishes only when $n_a = n_b$. The decay of the observed oscillations is governed by the radiative and non-radiative decay of the population of the states, and by dephasing.

With respect to the measurement method, the pulse propagation effect through a gain medium of finite length has to be included in the simulation. Supplementary Fig. 2c shows the single-cycle pulse after 2 mm of propagation in the modelled gain medium. The corresponding spectral content of the pulse is shown in Supplementary Fig. 2c. The response consists of two components—a resonant main part corresponding to the eigenfrequency of the two-level system, and a broadband part that is due to the instant response of the two-level system to the driving electric field. In the latter case, the population of the states is coherently driven by the electric field of the THz pulse and a partial Rabi oscillation occurs.

LETTERS

Observation of ferrotoroidic domains

Bas B. Van Aken[1,2], Jean-Pierre Rivera[3], Hans Schmid[3] & Manfred Fiebig[1,2]

Domains are of unparalleled technological importance as they are used for information storage and for electronic, magnetic and optical switches. They are an essential property of any ferroic material. Three forms of ferroic order are widely known: ferromagnetism, a spontaneous magnetization; ferroelectricity, a spontaneous polarization; and ferroelasticity, a spontaneous strain. It is currently debated whether to include an ordered arrangement of magnetic vortices as a fourth form of ferroic order, termed ferrotoroidicity. Although there are reasons to expect this form of order from the point of view of thermodynamics[1], a crucial hallmark of the ferroic state—that is, ferrotoroidic domains—has not hitherto been observed. Here ferrotoroidic domains are spatially resolved by optical second harmonic generation in $LiCoPO_4$, where they coexist with independent antiferromagnetic domains. Their space- and time-asymmetric nature relates ferrotoroidics to multiferroics with magnetoelectric phase control[2–5] and to other systems in which space and time asymmetry leads to possibilities for future applications.

A toroidic moment is generated by a vortex of magnetic moments, such as atomic spins or orbital currents[6]. Another example—a ring-shaped torus with an even number of current windings—is shown in Fig. 1a. It has been noted[6] that the toroidic moment is asymmetric under the reversal of time and under the reversal of space, a concept introduced in ref. 7. Suggestions to consider the spontaneous alignment of toroidic moments as a source for a fourth, space- and time-asymmetric form of ferroic order were originally made in refs 1, 8 and 9, and were discussed again recently[10,11]. In ferrotoroidic (FTO) materials, it is possible[6] to induce a magnetization **M** by an electric field **E** ($\mathbf{M}=\hat{\alpha}\mathbf{E}$) and a polarization **P** by a magnetic field **H** ($\mathbf{P}=\hat{\alpha}\mathbf{H}$), see Fig. 1b. This property (which is termed magnetoelectric effect[2,3] and expressed by the tensor $\hat{\alpha}$) explains some of the current interest in toroidic materials, because manipulation of magnetization by means other than magnetic fields has become an important technological challenge.

Any physical system can be characterized by its behaviour under a reversal of space and time. These so-called parity operations possess only two eigenvalues, −1 and +1, which correspond to a change of sign or no change of sign, respectively, when the parity operation is applied. Figure 2 shows the behaviour of all forms of ferroic order under space and time reversal. It is apparent that each of the four parity-group representations corresponds to a ferroic order[7,12], but only if ferrotoroidicity as ferroic order violating both space- and time-reversal symmetry is included. Space and time asymmetry relates FTO materials to (anti-)ferromagnetic ferroelectrics, called multiferroics. Multiferroics are space- and time-asymmetric because of the coexistence of two order parameters—one violating space-reversal symmetry and one violating time-reversal symmetry. They exhibit pronounced magnetoelectric correlations[13–15], and their symmetry relationship to the FTO order is another reason for the current interest in toroidic materials.

Attempts to demonstrate the presence of an FTO state as a fourth form of ferroic order were made by resonant X-ray diffraction[16]. It was further shown that ferrotoroidicity leads to an imbalance $\alpha_{ij} \neq \alpha_{ji}$ and to divergence of the magnetoelectric tensor components that is directly related to the presence of an FTO order parameter[9,17–19]. However, an essential property of any ferroic state—the existence of domains as regions with different and modifiable orientation of the corresponding order parameter—has not been demonstrated yet.

Because of the vortex nature of ferrotoroidicity, it is intrinsically linked to antiferromagnetism. (Similarly, many ferroelectrics are also ferroelastic.) To give a criterion to distinguish between the FTO state and the antiferromagnetic (AFM) state, we note that the former should display features that go beyond antiferromagnetism and lead to domains that are independent of the AFM domains. This can be the case in $LiCoPO_4$. It has the olivine crystal structure with *mmm* symmetry in the paramagnetic state. The Co^{2+} ions are located at positions like $(1/4+\varepsilon, 1/4, -\delta)$, where ε and δ are small displacements allowed by the *mmm* symmetry[20,21] (see Fig. 3). At 21.8 K, the Co^{2+} ions order in what was long assumed (on the basis of neutron

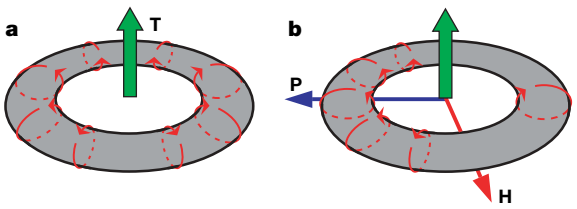


Figure 1 | Toroidic moment and magnetoelectric effect. a, Torus with an even number of current windings forming a toroidic moment, **T**. **b**, The magnetoelectric effect $\hat{\alpha}$ is illustrated by the current loops being shifted by the magnetic field (**H**), thus inducing an electric polarization (**P**). By rotating the figure by 90°, the asymmetry ($\alpha_{ij} = -\alpha_{ji}$) becomes obvious. Note that α_{ij} corresponds to a toroidal moment T_k ($i \neq j \neq k$).

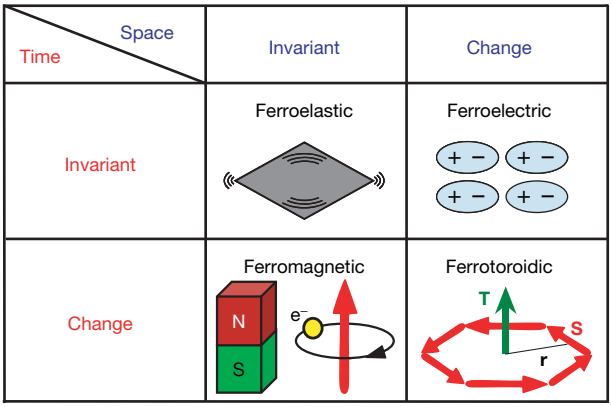


Figure 2 | All forms of ferroic order under the parity operations of space and time. For the toroidic moment, spins replace the current loops of Fig. 1.

[1]Max-Born-Institut, Max-Born-Straße 2A, 12489 Berlin, Germany. [2]HISKP, Universität Bonn, Nussallee 14-16, 53115 Bonn, Germany. [3]Department of Inorganic, Analytical and Applied Chemistry, University of Geneva, 30 quai Ernest-Ansermet, 1211 Geneva 4, Switzerland.

diffraction data[22]) to be a compensated AFM structure with spins along the y axis and mmm' symmetry. However, recently neutron diffraction revealed a weak (010) peak, which uniquely points to a rotation of the spins by $\varphi = 4.6^\circ$ away from the y axis and to a reduction of the symmetry to $2'$ with x as the twofold axis[21]. In Fig. 3 this spin structure is sketched. The spin part of the toroidic moment is described by $\mathbf{T} \propto \sum_n \mathbf{r}_n \times \mathbf{S}_n$ (see Fig. 2) with $\mathbf{r}_n$ the radius vector and $\mathbf{S}_n$ the spin of the n magnetic ions, taking the centre of the unit cell as the origin[1,8,10,11]. The cross-product is graphically expressed in Fig. 3b, where those spin components oriented perpendicular to $\mathbf{r}_n$ are shown for the x–z plane. Because $r_{1,3} > r_{2,4}$, the clockwise contribution of ions '1' and '3' is larger than the anticlockwise contribution of ions '2' and '4'. In total, this corresponds to a toroidic moment $T_y \propto \varepsilon$ pointing into the plane of Fig. 3b. Figure 3 reveals that the spin rotation by φ is essential for obtaining $T_y \neq 0$. Correspondingly, the rotation causes the reduction of symmetry to the group $2'$ in which $T_y \neq 0$ is permitted. Note that the $2'$ symmetry allows a magnetization along the spin direction. Weak ferromagnetism was indeed observed[21,23,24] but as Fig. 3 shows, its role in the generation of the toroidic moment is negligible.

How many different domains are expected for $LiCoPO_4$? AFM ordering reduces the symmetry from mmm to mmm' and the number of symmetry operations from sixteen to eight, corresponding to two AFM domains. Weak ferromagnetism along y reduces the symmetry from mmm' to $2'mm'$ and the number of symmetry operations from eight to four, corresponding to two ferromagnetic domains. However, the weak ferromagnetic component was shown to be coupled one-to-one to the AFM order parameter, so that this degree of freedom is lost[17]. Finally, spin rotation around x reduces the symmetry from $2'mm'$ to $2'$ and the number of symmetry operations from four to two, corresponding to two FTO domains. In principle, the low $2'$ symmetry may further lead to magnetically induced ferroelasticity or ferroelectricity[21]. However, we observed neither of these properties in polarization-optical and dielectric measurements on our samples (see Supplementary Information). So in summary, we expect to see two (weakly ferromagnetic) AFM and two FTO domains in $LiCoPO_4$.

A convenient technique for imaging (anti-) ferroic domain structures is optical second harmonic generation (SHG)[25]: an electromagnetic light field $\mathbf{E}$ at frequency ω is incident on a crystal, inducing a polarization $\mathbf{P}$ at frequency 2ω, which acts as source of an emitted, frequency-doubled light wave. This is expressed by $\mathbf{P}(2\omega) = \varepsilon_0 \hat{\chi} \mathbf{E}(\omega)\mathbf{E}(\omega)$ with $\hat{\chi}$ as SHG susceptibility. Following the Neumann principle, the symmetry determines the set of tensor components $\chi_{ijk} \neq 0$. Long-range ordering affects the symmetry. This leads to contributions $\propto \hat{\chi}(\mathcal{O})$ to the SHG, which couple, in first order, linearly to the order parameter $\mathcal{O}$ so that $\hat{\chi}(-\mathcal{O}) = -\hat{\chi}(\mathcal{O})$. Thus, SHG light from opposite domains has a 180° phase shift, which allows one to identify the domain structure[26]. As each form of (anti-) ferroic order affects symmetry in a different way, the SHG contributions from $\hat{\chi}(\mathcal{O}_1)$, $\hat{\chi}(\mathcal{O}_2)$, and so on in a multiferroic are in general differently polarized. This allows one to image even coexisting domain structures by polarization analysis[25].

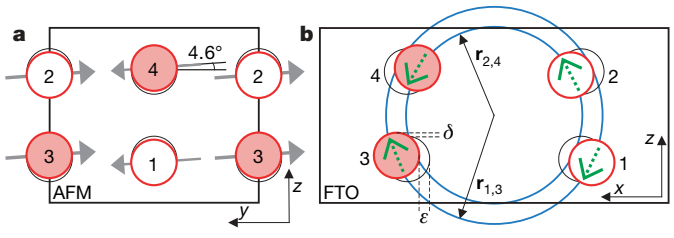


Figure 3 | Toroidic moment in $LiCoPO_4$. Co^{2+} ions at $x \approx 3/4$ (filled circles) and $x \approx 1/4$ (open circles), and shifts δ and ε, are shown. **a**, y–z plane and AFM order. Different thickness of the arrows indicates weak ferromagnetism along the spin direction. There are no spin spirals; all spins are (anti)parallel: $\mathbf{S}_n \| (0, \cos\varphi, \sin\varphi)$ with $\varphi = 4.6^\circ$ and $n = 1$ to 4. A change of sign of S_x or φ corresponds to a reversal of the AFM or FTO order parameter ($\pm\ell$ or $\pm T$), respectively. **b**, x–z plane and FTO order. Because $\mathbf{T} \propto \sum_n \mathbf{r}_n \times \mathbf{S}_n$, contributions to the toroidic moment T_y are only made by those components of $\mathbf{S}_n$ that are oriented perpendicular to $\mathbf{r}_n$. These spin components, magnified by 10 for clarity, are shown as green arrows.

Figure 4 shows the spectral dependence of the SHG intensity of $LiCoPO_4$ obtained with a set-up described elsewhere[25]. Tensor components $\chi_{ijk} \neq 0$ detected at 10 K on a polished (100)-oriented sample with thickness 122 μm are shown. Disappearance of the SHG intensity at the Néel temperature (T_N) confirms its coupling to the magnetic or the toroidic order parameter. The only symmetry that is compatible with the observed set of tensor components is $2'$ (see Supplementary Information). SHG thus confirms the proposed symmetry. In turn, this means that SHG couples to the toroidic moment T_y associated with this low symmetry.

Figure 5a shows an SHG image of $LiCoPO_4$ obtained with light from the χ_{yyz} component. The dark lines observed on a constant-intensity background are AFM domain walls. The aforementioned 180° phase shift between SHG light from opposite domains causes destructive interference at the walls, which leads to the dark lines[26] (the identification of AFM is discussed later). Figure 5c was obtained by interference of light from the χ_{yyz} and χ_{zyy} components, and displays a patchwork of bright and dark areas. Obviously two SHG contributions are present in Fig. 5c which interfere constructively (bright) or destructively (dark). A reversal of brightness occurs whenever an AFM domain wall is crossed because of the 180° phase shift occurring at the wall. But bright–dark interfaces are also present within the AFM domains. This indicates the presence of an additional type of domain with a distribution independent of the AFM domains. These are the FTO domains, and the red lines in Fig. 5c indicate the FTO domain walls.

At most photon energies the SHG images tend to show the domain structure of Fig. 5b, which indicates that the FTO contribution to the SHG is usually much smaller than the AFM contribution. This is reasonable, because the spin rotation leading to the toroidic moment is also small (only 4.6°, see Fig. 3). The spectral dependence of the SHG intensity thus confirms our assignment of order parameters to the domain structures. In order to observe both domain structures by interference, the AFM and the FTO contributions to the SHG must be sufficiently similar in amplitude and phase; this determines the photon energy chosen for Figs 5c and 6b, c.

Temperature cycles below T_N lead to a change of the FTO, but not of the AFM, domain structure (see, for example, Fig. 5c). This reflects the fact that the walls and energies of the two domain states are different, thereby confirming their different nature. Controlled

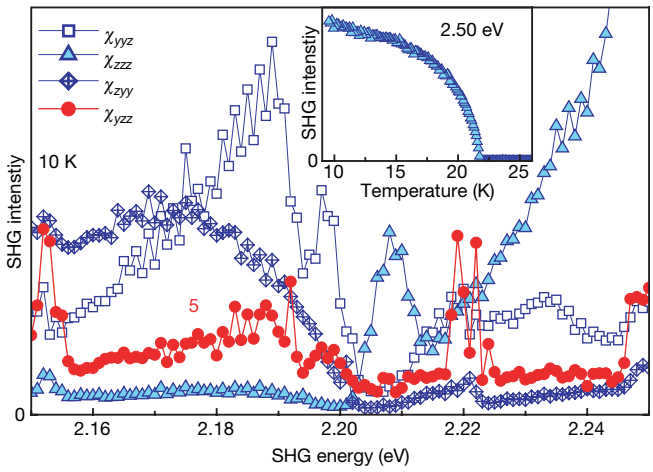


Figure 4 | SHG spectrum of $LiCoPO_4$(100). Contributions from independent non-zero tensor components χ_{ijk} are shown. The inset shows the temperature dependence. The SHG intensity drops to zero at 21.8 K, in excellent agreement with the ordering temperature from literature[17].

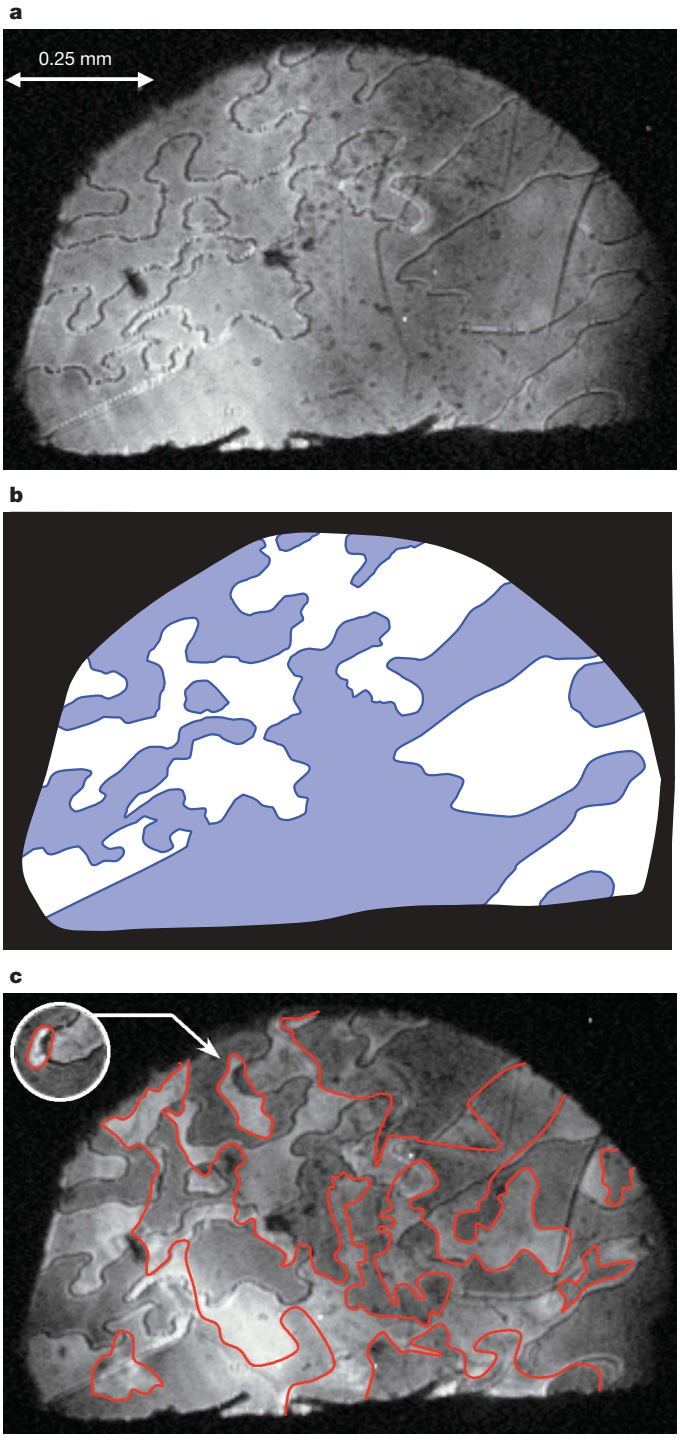


Figure 5 | Coexisting AFM and FTO domains of a $LiCoPO_4$(100) sample at 10 K imaged with SHG light at 2.197 eV. **a**, Obtained using light from χ_{yyz}. Dark lines are the AFM domain walls. **b**, Distribution of the AFM domains in **a**. **c**, Obtained using interfering light from χ_{yyz} and χ_{zyy}. Bright and dark areas are caused by the interference of AFM and FTO contributions to SHG (see text). Red lines indicate the FTO domain walls. Inset, FTO domain movement caused by a temperature cycle below T_N.

switching of the FTO state by setting a preferred direction of $\mathbf{T}$ with a toroidic field (for example, crossed magnetic and electric fields or a magnetic field with $\nabla \times \mathbf{H} \neq 0$)[1,10] is a more rigorous demonstration of ferroic switching, but is experimentally not compatible with our optical detection technique. However, we consider the observation of uniform regions of opposite and modifiable toroidic orientation as such strong evidence that an interpretation in terms of something other than FTO domains can be excluded.

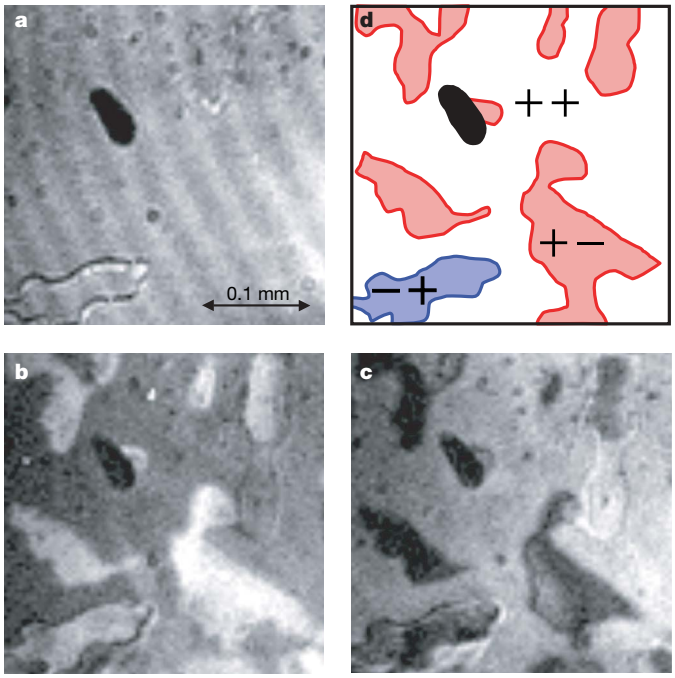


Figure 6 | Toroidic domains in a nearly single-AFM-domain $LiCoPO_4$(100) sample at 10 K. **a**, Obtained using SHG light from χ_{zzz} at 2.25 eV. The dark line at lower left is a single AFM domain wall. **b**, Obtained from $\chi_{yyz}+\chi_{zyy}$; **c**, obtained from $\chi_{yyz}-\chi_{zyy}$. **d**, Sketch of the AFM and the FTO domains. Relative to the large domain (AFM, $+\ell$; FTO, $+T$: shown as '++'), the red domains have $(+\ell,-T)$ and the blue domain has $(-\ell,+T)$. Black patch in all images, sample damage.

Slow temperature cycling produced the nearly single-domain AFM state shown in Fig. 6a. With exception of the bottom left corner, the domains in Fig. 6b are therefore of the FTO type. A rotation of the detected polarization of the SHG light around x by 90° leads to a reversal of brightness of all domains (Fig. 6c). The rotation exchanges (y, z) for $(y, -z)$. This changes the sign of the χ_{zyy} contribution, which inverts the interference and, thus, the contrast between the χ_{zyy} and the χ_{yyz} contribution. Note that contrast reversal is only possible if the SHG contributions responsible for the interference stem from independent sources, which are the AFM and the FTO domains shown in Fig. 6d.

Thus we have observed FTO domains in $LiCoPO_4$, where they coexist with independent AFM domains. Both types of domain were observed by spatially resolved optical SHG, and were identified on the basis of their symmetry properties and the crystal and spin structure of the compound. The observation of FTO domains strongly supports the concept of ferrotoroidicity as a fourth form of ferroic order along with ferromagnetism, ferroelectricity and ferroelasticity.

Ferroic order based on the simultaneous violation of space- and time-reversal symmetry by one or two order parameters is highly relevant for a broad variety of contemporary systems in which such violation leads to exciting possibilities for future applications. Examples are spin currents in spiral magnets leading to multiferroics with magnetoelectric phase control[2–5,13,14], polar superconductors[27] and trilayer superstructures with unidirectional light propagation[28].

Received 5 June; accepted 27 July 2007.

1. Dubovik, V. M. & Tugushev, V. V. Toroid moments in electrodynamics and solid-state physics. *Phys. Rep.* **187**, 145–202 (1990).
2. Fiebig, M. Revival of the magnetoelectric effect. *J. Phys. D* **38**, 123–152 (2005).
3. Eerenstein, W., Mathur, N. D. & Scott, J. F. Multiferroic and magnetoelectric materials. *Nature* **442**, 759–765 (2006).
4. Katsura, H., Nagaosa, N. & Balatsky, A. V. Spin current and magnetoelectric effect in noncollinear magnets. *Phys. Rev. Lett.* **95**, 057205 (2005).
5. Mostovoy, M. Ferroelectricity in spiral magnets. *Phys. Rev. Lett.* **96**, 067601 (2006).
6. Ginzburg, V. L., Gorbatsevich, A. A., Kopayev, Y. V. & Volkov, B. A. On the problem of superdiamagnetism. *Solid State Commun.* **50**, 339–343 (1984).

7. Ascher, E. Some properties of spontaneous currents. *Helv. Phys. Acta* **39,** 40–48 (1966).
8. Gorbatsevich, A. A. & Kopaev, Y. V. Toroidal order in crystals. *Ferroelectrics* **161,** 321–334 (1994).
9. Sannikov, D. G. Ferrotoroic phase transition in boracites. *Ferroelectrics* **219,** 177–181 (1998).
10. Schmid, H. On ferrotoroidics and electrotoroidic, magnetotoroidic and piezotoroidic effects. *Ferroelectrics* **252,** 41–50 (2001).
11. Ederer, C. & Spaldin, N. A. Towards a modern theory of toroidal moments in bulk periodic crystals. Preprint at ⟨http://xxx.lanl.gov/abs/0706.1974v1⟩ (2007).
12. Ascher, E. Kineto-electric and kinetomagnetic effects in crystals. *Int. J. Magn.* **5,** 287–295 (1974).
13. Kimura, T. *et al.* Magnetic control of ferroelectric polarization. *Nature* **426,** 55–58 (2003).
14. Hur, N., Sharma, P. A., Ahn, J. S., Guha, S. & Cheong, S.-W. Electric polarisation reversal and memory in a multiferroic material induced by magnetic fields. *Nature* **429,** 392–395 (2004).
15. Lottermoser, T. *et al.* Magnetic phase control by an electric field. *Nature* **430,** 541–544 (2004).
16. Arima, T. *et al.* Resonant magnetoelectric x-ray scattering in $GaFeO_3$: observation of ordering of toroidal moments. *J. Phys. Soc. Jpn* **74,** 1419–1422 (2005).
17. Kornev, I. *et al.* Magnetoelectric properties of $LiCoPO_4$ and $LiNiPO_4$. *Phys. Rev. B* **62,** 12247–12253 (2000).
18. Popov, Y. F. *et al.* Magnetoelectric effect and toroidal ordering in $Ga_{2-x}Fe_xO_3$. *J. Exp. Theor. Phys.* **87,** 146–151 (1998).
19. Ascher, E., Rieder, H., Schmid, H. & Stössel, H. Some properties of ferromagnetoelectric nickel-iodine borate, $Ni_3B_7O_{13}I$. *J. Appl. Phys.* **37,** 1404–1405 (1966).
20. Newnham, R. E. & Redman, M. J. Crystallographic data for $LiMgPO_4$, $LiCoPO_4$ and $LiNiPO_4$. *J. Am. Ceram. Soc.* **48,** 547 (1965).
21. Vaknin, D., Zarestky, J. L., Miller, L. L., Rivera, J.-P. & Schmid, H. Weakly coupled antiferromagnetic planes in single-crystal $LiCoPO_4$. *Phys. Rev. B* **65,** 224414 (2002).
22. Santoro, R. P., Segal, D. J. & Newnham, R. E. Magnetic properties of $LiCoPO_4$ and $LiNiPO_4$. *J. Phys. Chem. Solids* **27,** 1192–1193 (1966).
23. Kharchenko, N. F., Kharchenko, Y. N., Szymczak, R., Baran, M. & Schmid, H. Weak ferromagnetism in the antiferromagnetic magnetoelectric crystal $LiCoPO_4$. *Low Temp. Phys.* **27,** 895–898 (2001).
24. Rivera, J.-P. The linear magnetoelectric effect in $LiCoPO_4$ revisited. *Ferroelectrics* **161,** 147–164 (1994).
25. Fiebig, M., Pavlov, V. V. & Pisarev, R. V. Second-harmonic generation as a tool for studying electronic and magnetic structures of crystals: review. *J. Opt. Soc. Am. B* **22,** 96–118 (2005).
26. Fiebig, M., Lottermoser, T., Fröhlich, D. & Kallenbach, S. Phase resolved second harmonic imaging with nonideal laser sources. *Opt. Lett.* **29,** 41–44 (2004).
27. Bauer, E. *et al.* Heavy fermion superconductivity and magnetic order in noncentrosymmetric $CePt_3Si$. *Phys. Rev. Lett.* **92,** 027003 (2004).
28. Figotin, A. & Vitebskiy, I. Electromagnetic unidirectionality in magnetic photonic crystals. *Phys. Rev. B* **67,** 165210 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Collaborative Research Center (SFB) 608 and the Priority Program (SPP) 1133 of the DFG and the Swiss NSF for subsidy, and R. Boutellier and S. Gentil for help with growing the crystals. We further thank C. Ederer and N. A. Spaldin for many discussions. M.F. thanks T. Elsässer for continuous support. H.S. is indebted to E. Ascher for initiating him into the realm of symmetries and into the importance of the time-odd polar vector.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.F. (fiebig@hiskp.uni-bonn.de).

LETTERS

Isotopic evidence for Mesoarchaean anoxia and changing atmospheric sulphur chemistry

James Farquhar[1], Marc Peters[2], David T. Johnston[1], Harald Strauss[2], Andrew Masterson[1], Uwe Wiechert[3] & Alan J. Kaufman[1]

The evolution of the Earth's atmosphere is marked by a transition from an early atmosphere with very low oxygen content to one with an oxygen content within a few per cent of the present atmospheric level. Placing time constraints on this transition is of interest because it identifies the time when oxidative weathering became efficient, when ocean chemistry was transformed by delivery of oxygen and sulphate, and when a large part of Earth's ecology changed from anaerobic to aerobic[1]. The observation of non-mass-dependent sulphur isotope ratios in sedimentary rocks more than ~2.45 billion years (2.45 Gyr) old and the disappearance of this signal in younger sediments is taken as one of the strongest lines of evidence for the transition from an anoxic to an oxic atmosphere around 2.45 Gyr ago[1–5]. Detailed examination of the sulphur isotope record before 2.45 Gyr ago also reveals early and late periods of large amplitude non-mass-dependent signals bracketing an intervening period when the signal was attenuated[5–9]. Until recently, this record has been too sparse to allow interpretation, but collection of new data has prompted some workers[8] to argue that the Mesoarchaean interval (3.2–2.8 Gyr ago) lacks a non-mass-dependent signal, and records the effects of earlier and possibly permanent oxygenation of the Earth's atmosphere. Here we focus on the Mesoarchaean interval, and demonstrate preservation of a non-mass-dependent signal that differs from that of preceding and following periods in the Archaean. Our findings point to the persistence of an anoxic early atmosphere, and identify variability within the isotope record that suggests changes in pre-2.45-Gyr-ago atmospheric pathways for non-mass-dependent chemistry and in the ultraviolet transparency of an evolving early atmosphere.

The report seven years ago of non-mass-dependent isotope signals in a global distribution of Archaean and early Palaeoproterozoic sediments and metasediments[5] prompted a focused effort to understand the connections between the early sulphur cycle and the history of atmospheric oxygen. These non-mass-dependent effects do not follow the mass-dependent relationships $\delta^{33}S \approx 0.5\delta^{34}S$ and $\delta^{36}S \approx 2\delta^{34}S$, where the terms 0.5 and 2 reflect the relative mass differences between the numerator and denominator of $^{33}S/^{32}S$ and $^{36}S/^{32}S$ compared to those of $^{34}S/^{32}S$ (see Methods for definitions of notation and a more detailed description of non-mass-dependent and mass-dependent isotope effects). Work on this topic has yielded a growing data set of $\Delta^{33}S$ and $\Delta^{36}S$ (a measure of the deviation from a reference array that approximates low temperature chemical equilibrium and passes through bulk Earth composition; see Methods) that is fully complementary to the existing $\delta^{34}S$ data and interpretations made from this time series record. These data ($\delta^{34}S$, $\Delta^{33}S$ and $\Delta^{36}S$) provide new insights into the processes and conditions necessary to produce non-mass-dependent sulphur isotope effects in the Archaean and earliest Palaeoproterozoic, as well as into the post-2.45-Gyr-ago world, where mass-dependent isotope effects are preserved[7,10–12], and non-mass-dependent effects are largely absent—the exception being effects reported for some atmospheric samples and accumulations in ice cores[13–15].

Figure 1 illustrates the dramatic change in the magnitude of $\Delta^{33}S$ that occurred between ~2.45 and 2.32 Gyr ago[1,6,12]. The disappearance of large $\Delta^{33}S$ in the rocks younger than ~2.45 Gyr reflects the suppression of the non-mass-dependent effects. Concomitant enhancement of mass-dependent signals (brought about by significantly larger sources of sulphate sulphur from oxidative weathering) reflects a shift from a sulphur cycle conducive to preservation of atmospheric isotope effects to a sulphur cycle in which biological isotope effects dominate. A corresponding increase in the range of $\delta^{34}S$ (Fig. 2) has been attributed to the more significant role for sulphate reducing microorganisms in oceanic environments with higher sulphate concentrations. This figure also illustrates temporal variations in the Archaean $\Delta^{33}S$ record, including an apparent

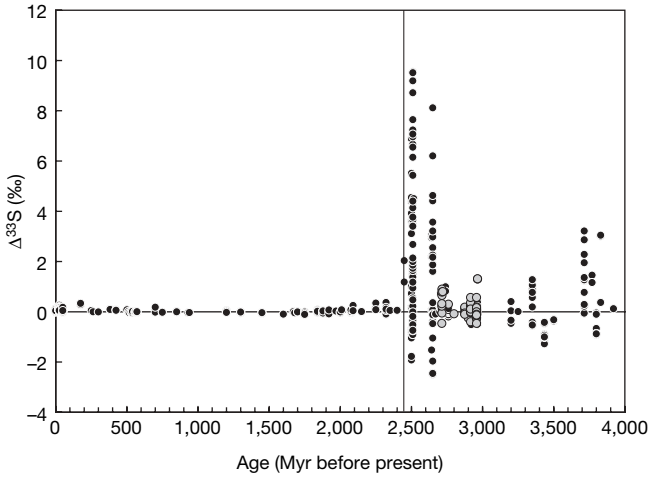


Figure 1 | Compilation of $\Delta^{33}S$ versus age for rock samples. Data are taken from refs 5–8, 12, 19–30. Ion probe data have been grouped by sample averages when measurements of different grains in the same sample did not vary outside reported uncertainties. The figure illustrates large $\Delta^{33}S$ before ~2.45 Gyr ago (indicated by vertical line), and small but measurable $\Delta^{33}S$ after 2.45 Gyr ago. Detailed examination of the pre-2.45-Gyr-ago record indicates an apparent minimum for $\Delta^{33}S$ during the Mesoarchaean (2.8–3.2 Gyr ago). Uncertainties are smaller than the symbol sizes for $\Delta^{33}S$. Filled black circles are data from literature; filled grey circles are data presented in this study.

[1]Department of Geology and ESSIC, University of Maryland, College Park, Maryland 20742, USA. [2]Geologisch-Paläontologisches Institut und Museum der Westfälischen Wilhelms-Universität Münster, Corrensstraße 24, 48149 Münster, Germany. [3]Department of Earth Sciences, Freie Universität Berlin, Malteserstr. 74-100, Haus B, C und N, 12249 Berlin, Germany.

minimum in the absolute range of $\Delta^{33}S$ in the Mesoarchaean (published and new data are presented in Supplementary Table 1). A question raised by the data, and one we focus on here, is whether this minimum reflects an absence of processes that produce non-mass-dependent fractionations[8], or whether it reflects something that caused the amplitude of $\Delta^{33}S$ to diminish, but did not turn off non-mass-dependent processes. The answer to this question rests with the cause and character of the non-mass-dependent effects.

The reports of non-mass-dependent sulphur isotope signals in modern atmospheric sulphate aerosols[13], and in sulphate produced from volcanogenic SO_2 recovered from horizons in Antarctic ice cores and firn[14,15], have been attributed to stratospheric reactions involving SO_2, and provide one link to the atmosphere[13–15]. Additional links to the atmosphere have been made on the basis of photolysis experiments with SO_2, a species inferred to be present in the atmosphere throughout Earth history. These experiments yield large non-mass-dependent isotope effects that exhibit a dependence on wavelength and are similar in many respects to the sulphur isotope compositions observed in the geologic record[2].

The connection to these photolysis experiments and to modern stratospheric chemistry is taken as evidence that the early atmosphere was (partially or wholly) transparent to deep ultraviolet radiation—to allow similar chemical reactions to occur. An atmosphere transparent to deep ultraviolet would imply a low column depth for ozone, and also for oxygen because ozone is maintained at present high levels by chemistry involving atmospheric oxygen. Models of atmospheric chemistry imply a second connection to oxygen concentrations in the atmosphere. This is because the pathways for the transfer of non-mass-dependent sulphur to Earth's surface as aerosols (S_8 and sulphate) depend on whether oxygen and other oxygen species were high enough to consume neutral sulphur species (for example, S, S_2 and S_3)[3]. In low oxygen conditions, S_8 aerosols are stabilized in addition to sulphate aerosols and this allows for a more efficient transfer of atmospheric non-mass-dependent signals to surface sulphur and sulphate pools.

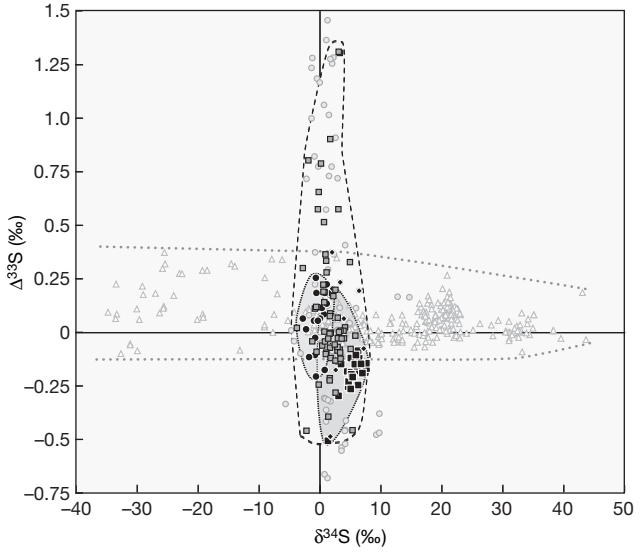


Figure 2 | Plot of $\Delta^{33}S$ versus $\delta^{34}S$ that illustrates the fields defined by pre- and post-2.45-Gyr-ago samples. The dotted grey line marks the limits of the observed post-2.45-Gyr-ago field. Fields are also included for published data, including samples from the Mozaan group[7] (filled black diamonds), samples from the Hardey Formation[8] (filled black circles), samples from the Mosquito Creek Formation[8] (filled black squares), and data from this study (filled grey squares). The field for the data from this study is outlined with a dashed black outline, as are the fields for literature data for the Hardey Formation, the Mosquito Creek Formation and the Mozaan Group, which are outlined with dotted black outlines and also shaded in grey. Grey unfilled triangles are literature data for post 2.45-Gyr-ago samples and grey filled circles circles are literature data for pre 2.45-Gyr-ago samples, excluding those listed above. Uncertainties for the data presented in this study are smaller than the symbol sizes for $\Delta^{33}S$ and $\delta^{34}S$.

To investigate further the possibility of a non-mass-dependent effect in the Mesoarchaean (and slightly post-Mesoarchaean) record, we illustrate the covariation of $\delta^{34}S$ with $\Delta^{33}S$ for Mesoarchaean samples (Fig. 2), which provides a means to explore the significance of the attenuated $\Delta^{33}S$ in this interval. Although the absolute magnitude of the $\Delta^{33}S$ effect for these samples is smaller than observed in other portions of the Archaean (and in particular, the earliest Palaeoproterozoic), the $\delta^{34}S$ versus $\Delta^{33}S$ field is entirely consistent with the overall field for the Archaean/earliest Palaeoproterozoic data; it is also distinct from samples younger than 2.45 Gyr when the atmosphere was undeniably oxidized. We include data presented by Ohmoto *et al.*[8] and Ono *et al.*[7] in Fig. 2 to illustrate the striking consistency of each existing set of Mesoarchaean data with the non-mass-dependent (or pre-2.45-Gyr ago) array rather than the mass-dependent (or post-2.45-Gyr ago) array.

Although the non-mass-dependent signal may have been attenuated during the Mesoarchaean, our compilation illustrates a larger range of $\Delta^{33}S$ variation than observed by Ohmoto *et al.*[8] and further suggests active non-mass-dependent chemistry. To support this contention, we provide an additional (and established) metric for distinguishing non-mass-dependent from mass-dependent signals with the concurrent measurement of ^{36}S abundances and the construction of $\Delta^{36}S/\Delta^{33}S$ arrays (Fig. 3)[5,7]. The low $\Delta^{36}S$ values in Fig. 3a reflect Rayleigh amplification of mass-dependent fractionation effects and define the limits of the mass-dependent field; they are well explained by theoretical considerations[16]. Covariation of $\Delta^{36}S$ with $\Delta^{33}S$ forms one of the principal arguments for a gas-phase photolytic origin of the effects observed in the Archaean/earliest Palaeoproterozoic record, and mass-dependent samples define a restricted field on a plot of $\Delta^{36}S$ versus $\Delta^{33}S$ that is distinct from the field defined by Archaean/earliest Palaeoproterozoic samples.

The Mesoarchaean data from this study, and one data point previously suggested to be anomalous by Ono *et al.*[7], define fields that are distinct from the greater Archaean data set and the mass-dependent post-2.45-Gyr intervals. The differences between the Mesoarchaean data set and the data set for older Archaean samples are inconsistent with a non-mass-dependent signal that was inherited from pre-existing terranes, and support the view of a pervasive anoxic atmosphere throughout the Archaean eon and into the earliest Palaeoproterozoic. Although some Mesoarchaean samples (for example, from the Nsuzi, Moodies and Mozaan[7] Groups) overlap the mass-dependent array for $\Delta^{36}S/\Delta^{33}S$, evidence for the large $\delta^{34}S$ range that is required to produce these minor isotope variations by mass-dependent processes (see, for example, ref. 16) is not observed in samples from this time interval. We also point out that the Mesoarchaean data form distinct fields when grouped at the formation level (Fig. 3b), illustrating fundamental changes in Mesoarchaean non-mass-dependent signals, and we highlight similarities in the Mesoarchaean signals from sequences in South Africa and those in Western Australia.

Insight into the origin of the Mesoarchaean isotope variation (a diminished $\Delta^{33}S$ amplitude and variable $\Delta^{36}S/\Delta^{33}S$) is provided by considering results from photochemical experiments and by comparisons with other Archaean intervals. The $\delta^{34}S$, $\Delta^{33}S$ and $\Delta^{36}S$ variations for some parts of the Archaean record—for example, the ~3.4-Gyr-old successions in Western Australia and South Africa—appear to be well-matched by experiments undertaken with ~193 nm radiation[2], and this observation led to the suggestion that deep-ultraviolet photolysis of sulphur dioxide provided an explanation for the origin of the non-mass-dependent signal in the Archaean eon. A subsequent study[6] documented a different relationship for $\delta^{34}S$ and $\Delta^{33}S$ variations (for the ~2.5-Gyr-old Mt McRae Shale), which fell between arrays produced by photolysis at 193 nm and experiments undertaken at longer wavelengths[17]: this led to a revised explanation for the non-mass-dependent signal—that it reflects a

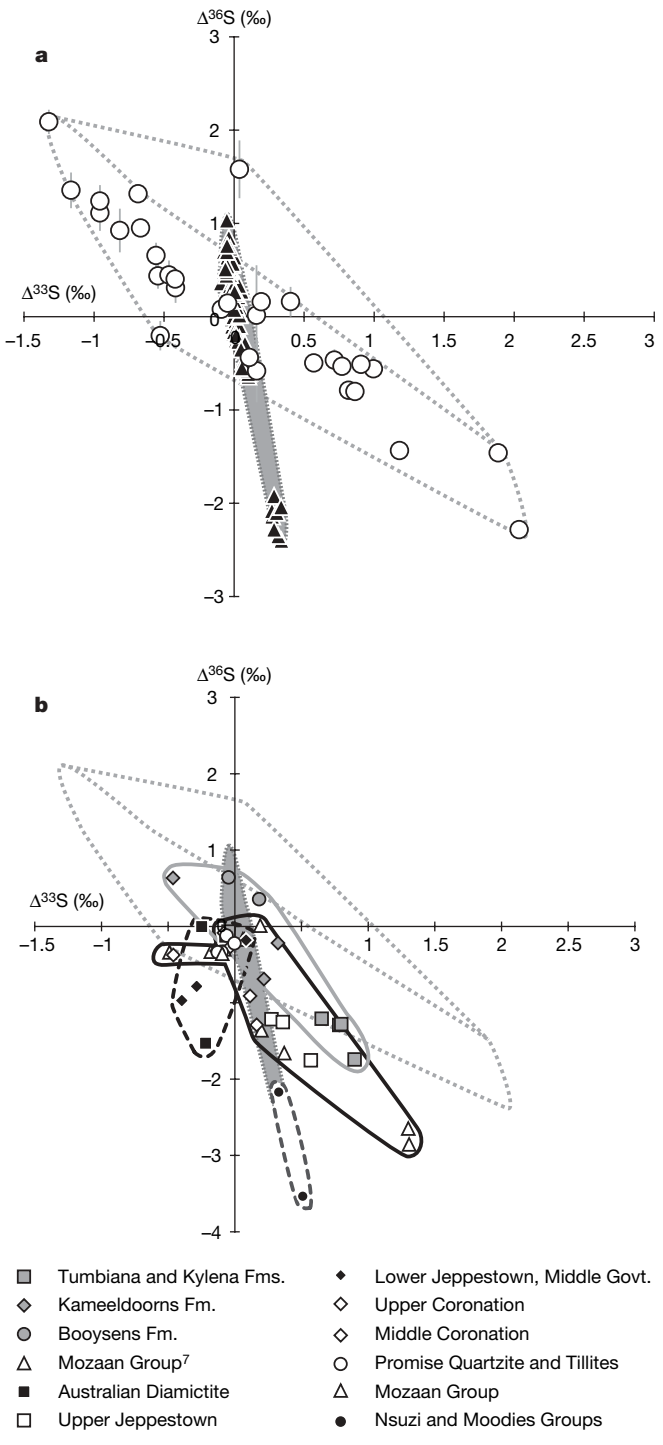


Figure 3 | Plot of Δ^{36}S versus Δ^{33}S for mass-dependent and mass-independent data. a, Plot of Δ^{36}S versus Δ^{33}S for pre-2.45-Gyr-ago samples (circles)[5] and for mass-dependent data (triangles) [16,19,20] (note that we have not included data from ref. 5 in the mass-dependent data set because of concerns about the exponent for Δ^{36}S noted in ref. 21). **b**, Plot of Δ^{36}S versus Δ^{33}S for samples reported in this study and by Ono *et al.*[7]. The field defined by the Archaean/earliest Palaeoproterozoic samples is outlined by a dotted line, and the field defined by the mass-dependent samples is outlined by a dotted line and shaded light grey. Data grouped into the field outlined by the solid grey line include data from the Kameeldoorns Formation (filled grey diamonds), the Tumbiana and Kylena Formations (filled grey squares) and the Booysens Formation (filled grey circles). Data grouped into the field outlined by a solid black line includes the Mozaan Group (unfilled triangles), the Upper Jeppestown Formation (unfilled squares), the Middle and Upper Coronation Shale (unfilled diamonds) and the Promise Quartzite and Tillites (unfilled circles). Data for the Lower Jeppestown and Middle Government subgroup (filled black diamonds) and from the basal Fortescue group associated with diamictites (filled black squares) are grouped separately in a field surrounded by a dashed black line, and data for samples from the Nsuzi Group and the Moodies Group (filled black circles) are in a field outlined by a dashed grey line. Uncertainties for Δ^{33}S are smaller than the symbol size, and uncertainties for Δ^{36}S are twice the symbol size.

combination of effects produced at shorter and longer ultraviolet wavelengths[6,18]. The data for the Mesoarchaean produce another relationship between δ^{34}S, Δ^{33}S and Δ^{36}S that adds complexity but may also provide a way to reconcile the observations made of Palaeoarchean[2] and Neoarchaean[6] samples if they reflect changes in the source reactions that occurred through Archaean time.

The photolysis experiments[2,18] exhibit a wavelength dependence that can account for the observed δ^{34}S, Δ^{33}S and Δ^{36}S relationships, and we speculate that changing transparency of the atmosphere may account for the observed minor sulphur isotope variation across the Archaean eon. In this context, the presence of variability for the Mesoarchaean data on plots of the Δ^{36}S versus Δ^{33}S (Fig. 3b) may also signal changes in atmospheric sulphur chemistry or transport. It is not clear whether the distinct arrays for samples grouped at the formation level within the Mesoarchaean indicate that the non-mass-dependent reactions tell us about specific states of the atmosphere, but we suggest that the fanning of these arrays rather than a coherence from formation to formation points to a control that varies on timescales represented by the formations themselves. Our data therefore point to continued low oxygen conditions throughout the Archaean, and imply changes in chemistry. These changes might be related to variations in the relative abundance of trace gas species (maybe SO_2), or to fluctuations in ultraviolet radiation transmitted through the atmosphere. Such fluctuations could reflect changes in atmospheric composition or the formation of atmospheric organic aerosol hazes.

METHODS SUMMARY

Sulphur isotope analyses were undertaken using standard fluorination methods and dual inlet gas-source mass spectrometry of sulphur extracted from samples by standard techniques (see Methods for details). Sulphur isotope ratios were measured for SF_6 as the analyte using a ThermoFinnigan MAT 253. The data are described using standard notation: $\delta^{34}S = 1{,}000 \times ((^{34}S/^{32}S)_{sample}/(^{34}S/^{32}S)_{ref} - 1)$, $\Delta^{33}S = 1{,}000 \times ((^{33}S/^{32}S)_{sample}/(^{33}S/^{32}S)_{ref} - ((^{34}S/^{32}S)_{sample}/(^{34}S/^{32}S)_{ref})^{0.515})$, and $\Delta^{36}S = 1{,}000 \times ((^{36}S/^{32}S)_{sample}/(^{36}S/^{32}S)_{ref} - ((^{34}S/^{32}S)_{sample}/(^{34}S/^{32}S)_{ref})^{1.9})$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 March; accepted 14 August 2007.

1. Holland, H. D. The oxygenation of the atmosphere and oceans. *Phil. Trans. R. Soc. B* **361**, 903–915 (2006).
2. Farquhar, J., Savarino, J., Airieau, S. & Thiemens, M. H. Observation of wavelength-sensitive mass-independent sulfur isotope effects during SO_2 photolysis: Implications for the early atmosphere. *J. Geophys. Res. E* **106**, 32829–32839 (2001).
3. Pavlov, A. A. & Kasting, J. F. Mass-independent fractionation of sulfur isotopes in Archean sediments: Strong evidence for an anoxic Archean atmosphere. *Astrobiology* **2**, 27–41 (2002).
4. Zahnle, K., Claire, M. & Catling, D. The loss of mass-independent fractionation in sulfur due to a Palaeoproterozoic collapse of atmospheric methane. *Geobiology* **4**, 271–283 (2006).
5. Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
6. Ono, S. *et al.* New insights into Archean sulfur cycle from mass-independent sulfur isotope records from the Hamersley Basin, Australia. *Earth Planet. Sci. Lett.* **213**, 15–30 (2003).
7. Ono, S., Beukes, N. J., Rumble, D. & Fogel, M. L. Early evolution of atmospheric oxygen from multiple sulfur and carbon isotope records of the 2.9 Ga Mozaan Group of the Pongola Supergroup, Southern Africa. *South Afr. J. Geol.* **109**, 97–108 (2006).
8. Ohmoto, H., Watanabe, Y., Ikemi, H., Poulson, S. R. & Taylor, B. E. Sulphur isotope evidence for an oxic Archaean atmosphere. *Nature* **442**, 908–911 (2006).
9. Kasting, J. F. & Ono, S. Palaeoclimates: the first two billion years. *Phil. Trans. R. Soc. B* **361**, 917–929 (2006).

10. Farquhar, J. *et al.* Multiple sulfur isotopic interpretations of biosynthetic pathways. *Geobiology* **1,** 27–36 (2003).
11. Johnston, D. T. *et al.* Active microbial sulfur disproportionation in the Mesoproterozoic. *Science* **310,** 1477–1479 (2005).
12. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427,** 117–120 (2003).
13. Romero, A. B. & Thiemens, M. H. Mass-independent sulfur isotopic compositions in present-day sulfate aerosols. *J. Geophys. Res. D* **108,** doi:10.1029/2003JD003660 (2003).
14. Savarino, J., Romero, A., Cole-Dai, J., Bekki, S. & Thiemens, M. H. UV induced mass-independent sulfur isotope fractionation in stratospheric volcanic sulfate. *Geophys. Res. Lett.* **30,** doi:10.1029/2003GL018134 (2003).
15. Baroni, M., Thiemens, M. H., Delmas, R. J. & Savarino, J. Mass-independent sulfur isotopic compositions in stratospheric volcanic eruptions. *Science* **315,** 84–87 (2007).
16. Ono, S., Wing, B., Johnston, D., Farquhar, J. & Rumble, D. Mass-dependent fractionation of quadruple stable sulfur isotope system as a new tracer of sulfur biogeochemical cycles. *Geochim. Cosmochim. Acta* **70,** 2238–2252 (2006).
17. Farquhar, J. & Wing, B. A. Multiple sulfur isotopes and the evolution of the atmosphere. *Earth Planet. Sci. Lett.* **213,** 1–13 (2003).
18. Farquhar, J., Savarino, J., Jackson, T. L. & Thiemens, M. H. Evidence of atmospheric sulphur in the martian regolith from sulphur isotopes in meteorites. *Nature* **404,** 50–52 (2000).
19. Ono, S., Shanks, W. C., Rouxel, O. J. & Rumble, D. ^{33}S constraints on the seawater sulfate contribution in modern seafloor hydrothermal vent sulfides. *Geochim. Cosmochim. Acta* **71,** 1170–1182 (2007).
20. Johnston, D. T. *et al.* Evolution of the oceanic sulfur cycle at the end of the Paleoproterozoic. *Geochim. Cosmochim. Acta* **70,** 5723–5739 (2006).
21. Farquhar, J., Bao, H. M., Thiemens, M. H., Hu, G. X. & Rumble, D. Questions regarding Precambrian sulfur isotope fractionation — Response. *Science* **292,** U6–U7 (2001).
22. Farquhar, J. & Wing, B. A. in *Mineral Deposits and Earth Evolution* (eds McDonald, I., Boyce, A. J., Butler, I. B., Herrington, R. J. & Polya, D. A.) 167–177 (Spec. Publ. 248, Geological Society, London, 2005).
23. Whitehouse, M. J. *et al.* Integrated Pb- and S-isotope investigation of sulphide minerals from the early Archaean of southwest Greenland. *Chem. Geol.* **222,** 112–131 (2005).
24. Papineau, D., Mojzsis, S. J., Coath, C. D., Karhu, J. A. & McKeegan, K. D. Multiple sulfur isotopes of sulfides from sediments in the aftermath of Paleoproterozoic glaciations. *Geochim. Cosmochim. Acta* **69,** 5033–5060 (2005).
25. Mojzsis, S. J., Coath, C. D., Greenwood, J. P., McKeegan, K. D. & Harrison, T. M. Mass-independent isotope effects in Archean (2.5 to 3.8 Ga) sedimentary sulfides determined by ion microprobe analysis. *Geochim. Cosmochim. Acta* **67,** 1635–1658 (2003).
26. Hu, G. X., Rumble, D. & Wang, P. L. An ultraviolet laser microprobe for the in situ analysis of multisulfur isotopes and its use in measuring Archean sulfur isotope mass-independent anomalies. *Geochim. Cosmochim. Acta* **67,** 3101–3118 (2003).
27. Ono, S., Wing, B., Johnston, D., Farquhar, J. & Rumble, D. Mass-dependent fractionation of quadruple stable sulfur isotope system as a new tracer of sulfur biogeochemical cycles. *Geochim. Cosmochim. Acta* **70,** 2238–2252 (2006).
28. Cates, N. L. & Mojzsis, S. J. Chemical and isotopic evidence for widespread Eoarchean metasedimentary enclaves in southern West Greenland. *Geochim. Cosmochim. Acta* **70,** 4229–4257 (2006).
29. Papineau, D. & Mojzsis, S. J. Mass-independent fractionation of sulfur isotopes in sulfides from the pre-3770 Ma Isua Supracrustal Belt, West Greenland. *Geobiology* **4,** 227–238 (2006).
30. Kamber, B. S. & Whitehouse, M. J. Micro-scale sulphur isotope evidence for sulphur cycling in the late Archean shallow ocean. *Geobiology* **5,** 5–17 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by funds from the NSF EAR, NASA NAI and NASA EXB progammes (to J.F.), and revisions were undertaken while J.F. was supported by a visiting appointment at the IPG of Paris. Other support for this work came from NSF (A.J.K.) and the DFG (H.S.). The manuscript was improved by reviews and comments from P. Knauth and H. Ohmoto. P. Cartigny is thanked for reading and commenting on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.F. (jfarquha@geol.umd.edu).

METHODS

We use a framework for distinguishing non-mass-dependent and mass-dependent isotope effects made using data collected from the geologic record that is anchored in an understanding of the way that mass-dependent processes operate. Our framework is distinct from those that place the distinction between non-mass-dependent and mass-dependent fractionations using only a threshold of $\Delta^{33}S$ (for example, at ~0.4‰; ref. 31) because it also considers information present in correlations between $\delta^{34}S$, $\Delta^{33}S$ and $\Delta^{36}S$. While this provides a more robust test of the presence of non-mass-dependent effects, we recognize that there will be an overlap in the fields for mass-dependent and non-mass-dependent signals. In some cases, samples that belong to non-mass-dependent populations may overlap with the range of the mass-dependent population.

In this study, sulphur from 66 samples were extracted using chromium reduction and converted to silver sulphide, converted to SF_6 by fluorination in a Ni reaction vessel at 250 °C with ~10× excess F_2 gas, and then purified by cryogenic and chromatographic techniques as described elsewhere[32]. Isotope ratios were determined by dual inlet IRMS using a ThermoFinnigan MAT 253 mass spectrometer with simultaneous monitoring of 127, 128, 129 and 131 (AMU/e^-). Uncertainties on mass-dependent reference materials are better than ±0.2‰, ±0.01‰ and ±0.2‰ in $\delta^{34}S$, $\Delta^{33}S$ and $\Delta^{36}S$, respectively. Uncertainties on the measurements reported here are estimated to be better than 0.2‰, ±0.02‰ and ±0.2‰. The results of our measurements are presented in Supplementary Table 1 and in Figs 1, 2 and 3.

Mass-dependent and non-mass-dependent fractionation. Geological interpretation of the sulphur isotope record has benefited significantly from the context provided by an understanding of the mechanisms that produce mass-dependent and non-mass-dependent isotopic fractionations. Mass-dependent fractionation effects arise when variations in chemical and physical properties of isotopologous species depend on the relative differences in mass of the atoms in a molecule, or the mass of the molecular species itself. Mass-dependent fractionations include, but are not limited to, equilibrium isotope exchange, diffusion, gravitational separation and biological effects[32–36]. Non-mass-dependent fractionations arise when variations in chemical and physical properties of molecular species depend on factors other than or in addition to the mass of the constituent isotopes[37–39]. Non-mass-dependent fractionations include those associated with symmetry changes of molecular species that arise from a particular isotope substitution, effects associated with internal conversions and intersystem crossings between bound and unbound states, effects associated with nuclear and electronic spin coupling, effects associated with self or mutual shielding, and effects that result from the physical production or destruction of an isotope. It has been established that non-mass-dependent and mass-dependent effects can occur together during chemical reactions or physical processes. For the purposes of this discussion, we will discuss non-mass-dependent effects as those that produce large deviations from mass-dependent fractionation arrays and involve factors other than mass, and we therefore do not include effects associated with mixing of different pools that are fractionated by mass-dependent processes. Our mass-dependent reference array is defined both by model predictions that take into consideration chemical and kinetic mass-dependent reactions and also by observations from oceanic sedimentary rocks. For this comparison, we have not included analyses of reagents or of samples of unknown provenance or age reported in refs 39 or 40.

We interpret the sulphur isotope data presented in Ohmoto *et al.*[31] to indicate that primary non-mass-dependent chemical effects were part of the sulphur cycles during deposition of the 2.76-Gyr-old Hardey Formation and the 2.92-Gyr-old Mosquito Creek Formation. These workers suggested that the signal observed in the rock record between 2.8 and 3.0 Gyr ago may reflect alternative processes, such as (1) high altitude chemistry involving sulphur dioxide photochemistry, (2) Rayleigh amplification of mass-dependent effects such as those that occur at fractionation crossovers, or (3) non-mass-dependent chemical reactions associated with the sulphate reduction reactions by amino acids. The first possibility has been discussed in ref. 41 and considered unlikely, considering the distribution of volcanogenic sulphur in the atmosphere and constraints imposed by models of volcanic plume ascent, and the second suggestion can be ruled out on the basis of relationships between $\Delta^{33}S$ and $\Delta^{36}S$ that do not form a single array as is expected for Rayleigh amplification of a mass-dependent process, even one occurring at fractionation crossovers. The third suggestion remains to be demonstrated in a form that can be evaluated in the context of the emerging record of the four sulphur isotopes that includes evidence for systematic changes of $\Delta^{36}S/\Delta^{33}S$, and possibly of systematic changes in the magnitude of the $\Delta^{33}S$ and the relationship between $\delta^{34}S$ and $\Delta^{33}S$. The viability of alternative hypotheses must be demonstrated to be consistent with the $\Delta^{36}S/\Delta^{33}S$ and $\delta^{34}S/\Delta^{33}S$. It also should account for any other observations in the record, including the possibility that there may be basinal and global correlations in the record and systematic variations in the relationships between these various measures of sulphur isotope composition.

Our conclusion primarily differs from those in Ohmoto *et al.*[31] because we have included constraints implied by the $\delta^{34}S$ data, which was not used to make the assertions in Ohmoto *et al.*[31]. This evaluation can be tested further using $\Delta^{36}S$ and the relationships between $\Delta^{36}S$ and $\Delta^{33}S$. For instance, it is possible that some of the $\Delta^{36}S/\Delta^{33}S$ variation in the Archaean data set may reflect the convolution of a mass-dependent signal with a non-mass-dependent signal of Archaean origin, and this possibly accounts for the variation in the $\Delta^{36}S$ at $\Delta^{33}S = 0$ for the larger Archaean data set, but this explanation does not account for the full magnitude of the $\Delta^{36}S$ variability in the Mesoarchaean and changes in $\Delta^{36}S/\Delta^{33}S$. We also recognize the possibility that processes other than photochemistry may contribute to the signals that are observed, but note that experimental evidence for non-mass-dependent effects involving gases of volcanogenic origin in combination with constraints from chemical models of anoxic atmospheres anticipates these effects in Earth's earliest anoxic environments. A subtle, but important point here is that the sulphur signal cannot be simply be an inherited detrital component if it has a different multiple sulphur isotope character, and even if there are detrital components, an authigenic Mesoarchaean component is inferred. An unresolved aspect of this signal in the Archaean sections is the observation of larger non-mass-dependent signals in shales, banded iron formations, and barites when compared to coarser clastic sediments.

31. Ohmoto, H., Watanabe, Y., Ikemi, H., Poulson, S. R. & Taylor, B. E. Sulphur isotope evidence for an oxic Archaean atmosphere. *Nature* **442,** 908–911 (2006).
32. Johnston, D. T. *et al.* Evolution of the oceanic sulfur cycle at the end of the Paleoproterozoic. *Geochim. Cosmochim. Acta* **70,** 5723–5739 (2006).
33. Farquhar, J. *et al.* Multiple sulfur isotopic interpretations of biosynthetic pathways *Geobiology* **1,** 27–36 (2003).
34. Johnston, D. T. *et al.* Active microbial sulfur disproportionation in the Mesoproterozoic. *Science* **310,** 1477–1479 (2005).
35. Ono, S., Shanks, W. C., Rouxel, O. J. & Rumble, D. ^{33}S constraints on the seawater sulfate contribution in modern seafloor hydrothermal vent sulfides. *Geochim. Cosmochim. Acta* **71,** 1170–1182 (2007).
36. Farquhar, J. & Wing, B. A. Multiple sulfur isotopes and the evolution of the atmosphere. *Earth Planet. Sci. Lett.* **213,** 1–13 (2003).
37. Thiemens, M. H. Atmosphere science — Mass-independent isotope effects in planetary atmospheres and the early solar system. *Science* **283,** 341–345 (1999).
38. Thiemens, M. H. History and applications of mass-independent isotope effects. *Annu. Rev. Earth Planet. Sci.* **34,** 217–262 (2006).
39. Hulston, J. R. & Thodem, H. G. Variations in S^{33}, S^{34} and S^{36} contents of meteorites and their relation to chemical and nuclear effects. *J. Geophys. Res.* **70,** 3475–3484 (1965).
40. Gao, X. & Thiemens, M. H. systematic study of sulfur isotopic composition in iron-meteorites and the occurrence of excess ^{33}S and ^{36}S. *Geochim. Cosmochim. Acta* **55,** 2671–2679 (1991).
41. Farquhar, J., Savarino, J., Airieau, S. & Thiemens, M. H. Observation of wavelength-sensitive mass-independent sulfur isotope effects during SO_2 photolysis: Implications for the early atmosphere. *J. Geophys. Res.* **106,** 32829–32839 (2001).

LETTERS

Attribution of observed surface humidity changes to human influence

Katharine M. Willett[1,2], Nathan P. Gillett[1], Philip D. Jones[1] & Peter W. Thorne[2]

Water vapour is the most important contributor to the natural greenhouse effect, and the amount of water vapour in the atmosphere is expected to increase under conditions of greenhouse-gas-induced warming, leading to a significant feedback on anthropogenic climate change[1–3]. Theoretical and modelling studies predict that relative humidity will remain approximately constant at the global scale as the climate warms, leading to an increase in specific humidity[1,4,5]. Although significant increases in surface specific humidity have been identified in several regions[6–9], and on the global scale in non-homogenized data[10], it has not been shown whether these changes are due to natural or human influences on climate. Here we use a new quality-controlled and homogenized gridded observational data set of surface humidity, with output from a coupled climate model, to identify and explore the causes of changes in surface specific humidity over the late twentieth century. We identify a significant global-scale increase in surface specific humidity that is attributable mainly to human influence. Specific humidity is found to have increased in response to rising temperatures, with relative humidity remaining approximately constant. These changes may have important implications, because atmospheric humidity is a key variable in determining the geographical distribution[11–13] and maximum intensity[14] of precipitation, the potential maximum intensity of tropical cyclones[15], and human heat stress[16], and has important effects on the biosphere[17] and surface hydrology[17,18].

We use a new gridded observational data set of surface specific humidity, HadCRUH[19], described in more detail in the Methods section. Global mean specific humidity, calculated from monthly anomalies over grid points with at least 75% of monthly means present, is shown by a continuous black line in Fig. 1 for the period 1973–2002. Over this period the observed global mean specific humidity in this data set increased by 0.07 g kg^{-1} per decade. This is similar to the 1976–2004 trend in global mean surface specific humidity of 0.06 g kg^{-1} per decade previously calculated from non-homogenized data[10]. The 1997–1999 mean is slightly higher than the 2000–2002 mean probably because of the large El Niño in 1997–1998, which strongly warmed the tropics and thereby substantially increased global mean specific humidity[19].

We compare observed variations in surface specific humidity with simulated variations from two four-member ensembles of the third Hadley Centre Coupled Model (HadCM3)[20] with anthropogenic forcings only (ANT) and natural forcings only (NAT)[21] over the period 1973–1999. This model exhibits realistic variability in global mean surface specific humidity (Supplementary Fig. 2). In the absence of archived surface specific humidity from the model's simulations with combined anthropogenic and natural forcing, we compare observed changes with the sum of the anthropogenic and natural responses. The summed anthropogenic and natural response shows an upward trend similar to that observed, and reproduces some of the interannual variability seen in the observations (Fig. 1, solid red line). We expect that such a linear sum represents a good approximation of the simulated response to combined anthropogenic and natural forcing, because the simulated temperature response to individual forcings is generally found to add linearly[22], and although the specific humidity q has an exponential dependence on temperature, a linear approximation is probably valid for the relatively small temperature anomalies considered here. Comparison with the separate ANT and NAT responses suggests that the observed upward trend is due mainly to anthropogenic forcing.

An increase in global mean specific humidity may reflect one of the following: a warming under conditions of constant relative humidity, an increase in relative humidity under conditions of constant temperature, a change in the geographical distribution of relative humidity under conditions of constant temperature, or some combination of these. To identify the changes in specific humidity directly associated with temperature changes, we use the gridded observational temperature data set HadCRUT3 (refs 23, 24) and an integrated form of the Clausius–Clapeyron equation (Methods) to calculate the specific humidity for each month and grid point, assuming constant climatological relative humidity and dry-air surface pressure. Temperature changes are found to explain the trend and variability well in both the observations (Fig. 1, dotted black line),

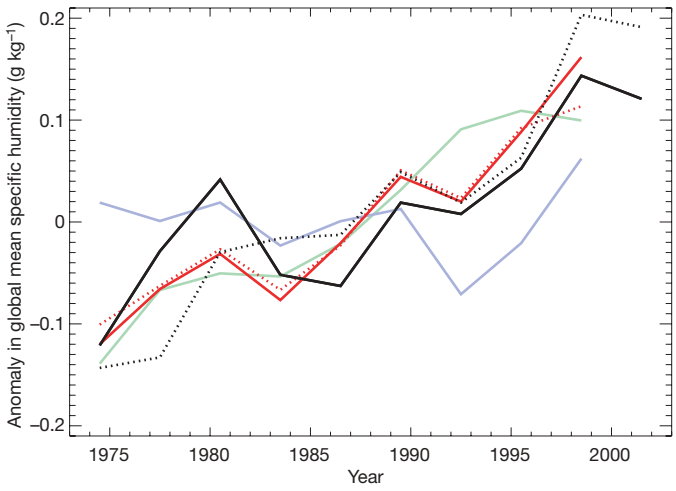


Figure 1 | Simulated and observed anomalies of global mean surface specific humidity. Non-overlapping three-year mean anomalies in global mean specific humidity in HadCRUH (solid black), and ensemble means of HadCM3 simulations of the response to natural forcing (blue), anthropogenic forcing (green), and the sum of the two (solid red). Dotted lines show changes in specific humidity due to temperature changes under conditions of constant relative humidity from observations (dotted black), and HadCM3 (dotted red).

[1]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [2]Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK.

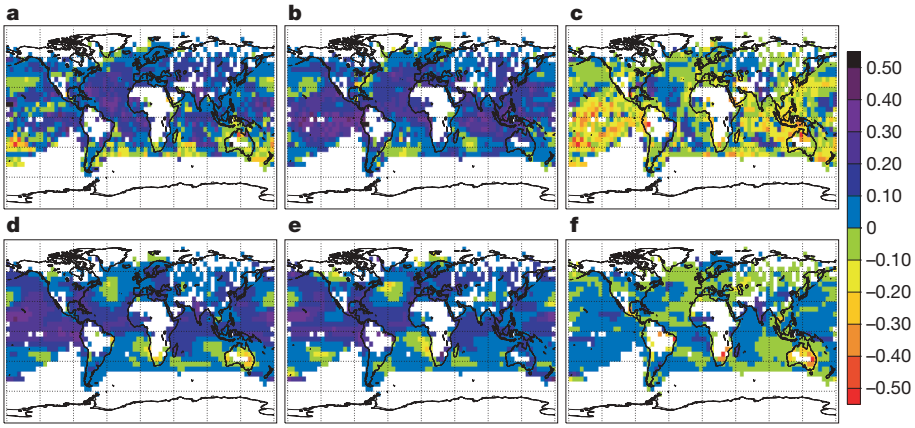


Figure 2 | Observed (top row) and simulated (bottom row) trends in specific humidity over the period 1973–1999, in g kg^{-1} per decade. Observed trends in q (**a**), and the sum of trends simulated in response to anthropogenic and natural forcing (**d**), are compared with trends calculated from observed (**b**) and simulated (**e**) temperature changes under the assumption of constant relative humidity; the residual (actual trend minus temperature-induced trend) is shown in **c** and **f**.

and the simulated ANT+NAT response (Fig. 1, dotted red line). Agreement is particularly close in the model, which may be explained partly by the fact that the simulated response is an ensemble mean. If we assume that specific humidity anomalies contain a temperature-forced component and an independent random component, we would expect the variance of the random residual to be one-quarter the size for a four-member ensemble mean, and half the size for the sum of two four-member ensemble means as we have here. On this basis, we would expect the average distance between solid and dotted black lines to be a factor $\sqrt{2}$ larger than the average distance between the solid and dotted red lines. The fact that this difference for the observations is more than a factor $\sqrt{2}$ larger than this difference for the model may be due in part to an apparent positive bias in marine humidity data before 1982 (Methods). Alternatively, temperature and specific humidity may be too closely correlated in the model[10]. Nonetheless the upward trend in q and the close correlation between q and temperature provide an additional piece of independent evidence corroborating the recent rapid warming seen in surface temperature.

The spatial pattern of specific humidity trends is also broadly consistent between the model and observations (Fig. 2a, d), and the largest increases in specific humidity are over the tropics. As a result of the approximately exponential dependence of saturation vapour pressure on temperature, a larger response of specific humidity is expected for a given change in temperature in warmer regions, under conditions of constant relative humidity. Consistent with this expectation is the observation that increases in q in the Northern Hemisphere are largest in the summer[19]. Increasing trends in specific humidity have been observed over most of the globe, with the main exceptions being Australia and parts of the Southern Ocean. The broad pattern of trends in specific humidity is found to be well explained by temperature trends (Fig. 2b, e). Residual changes in q are generally small in the model (Fig. 2f) but are predominantly negative in the observations (Fig. 2c). Decreases over much of the tropical oceans indicate a decrease in relative humidity in these regions: a consistent picture is obtained by examining relative humidity directly, and may relate in part to an apparent positive bias in marine humidity measurements before 1982 (Methods). Although simulated and observed residual trends seem consistent in some regions (for example, drying over Australia, moistening over India), the regression coefficient of the observed residual trends against the simulated residual trends is close to zero, indicating no detectable change in relative humidity[25].

To test objectively for the presence of an anthropogenic or natural response in observations of q, we used a detection and attribution analysis to compare simulated and observed changes[26,27]. After sampling the model in grid boxes in which monthly observed specific humidity anomalies were present, we took non-overlapping three-year means of simulated and observed anomalies in specific humidity over the common period of simulations and observations of 1973–1999 and projected onto the 25 lowest-order (T4) spherical harmonics to retain only the largest spatial scales in our analysis[26]. We regressed the observed spatio-temporal pattern of specific humidity anomalies against the simulated response to anthropogenic and natural forcings using a total least-squares optimal regression[27] and a projection of the simulated and observed spatio-temporal patterns onto the first 30 empirical orthogonal functions (EOFs) of the control simulation, although results were robust to variations in this EOF truncation (Supplementary Fig. 3). Internal variability was estimated from 1,200 years of HadCM3 control simulation: half was used to estimate EOFs, and half to assess uncertainties in the regression coefficients. Both ANT and NAT regression coefficients were found to be inconsistent with zero (Fig. 3), indicating a detectable response to both anthropogenic and natural forcing. The ANT regression coefficient was found to be consistent with one, and residual variability was found to be consistent with control variability, indicating

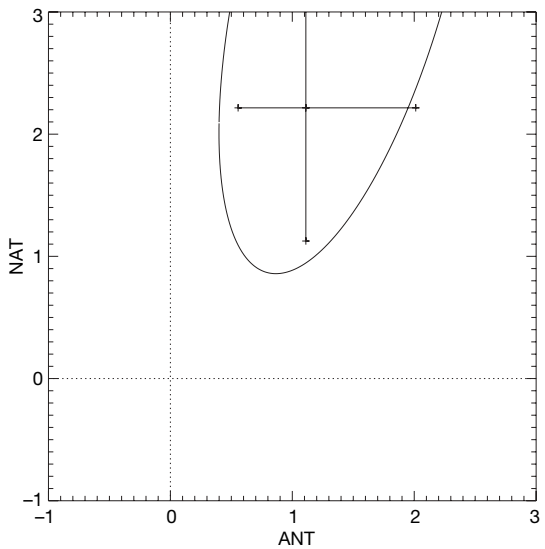


Figure 3 | Regression coefficients of observed specific humidity anomalies against the simulated response to anthropogenic (x axis) and natural (y axis) forcing. Results are based on nine three-year mean anomalies over the period 1973–1999 projected onto T4 spherical harmonics. One-dimensional 5–95% uncertainty ranges and a curve enclosing 90% of the estimated joint distribution are shown.

consistency between simulations and observations[28], although the NAT regression coefficient was found to be larger than one, suggesting that the natural response may be underestimated in the model. The natural response is probably dominated by volcanoes, which induce a global cooling for two to three years after major eruptions, which in turn decreases q (ref. 29). The response to both anthropogenic and natural forcing was detected in zonal means and T4 spherical harmonics over a large range of EOF truncations (Supplementary Fig. 3), and an anthropogenic response was also detectable in the global mean alone.

The results presented here add to an increasing body of evidence that atmospheric water vapour has exhibited a significant upward trend over recent decades[10,12,29,30]. Although such a moistening has long been predicted as a response to warming induced by greenhouse gases[1,4], this study demonstrates that the observed increase in surface specific humidity is directly attributable to anthropogenic influence and is distinct from the predicted response to natural forcing. The pattern of observed increases in specific humidity is largely consistent with the response to observed warming under conditions of constant relative humidity[10,25]. Recent results indicate that the precipitation response to anthropogenic forcing may be underestimated in climate models[12,13]; our demonstration of consistency between the simulated and observed specific humidity response to anthropogenic influence may be useful in identifying the reasons for this discrepancy and for constraining predictions of future hydrological change. Although radiative forcing of the climate is dominated by changes in the amount of water vapour in the upper troposphere[4], anthropogenic moistening of the surface and lower atmosphere is likely to have important implications for extreme precipitation, tropical cyclones and human heat stress.

METHODS SUMMARY

HadCRUH contains monthly mean q and relative humidity at $5^\circ \times 5^\circ$ resolution over the period 1973–2003 and is derived principally from land and marine measurements of dewpoint temperature. Observations are quality-controlled and homogenized before being included in the data set. There is some evidence of a positive bias in the marine specific humidity data before 1982 (refs 19, 25). Analysis suggests that this bias relates to changes in recording practice. Although we suggest that this issue should be further investigated when this data set is updated, we note that this bias would tend to reduce the calculated observed trend in q and would therefore lead to a conservative bias in the detection results presented here.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 March; accepted 30 August 2007.

1. Arrhenius, S. On the influence of carbonic acid in the air upon the temperature of the ground. *Phil. Mag.* (5) **41,** 237–276 (1896).
2. Hansen, J. *et al.* Climate sensitivity: Analysis of feedback mechanisms. *Geophys. Monogr.* **29,** 130–163 (1984).
3. Alley, R. *et al. Climate change 2007: The Physical Science Basis, Summary for Policymakers* (IPCC, Geneva, 2007).
4. Held, I. M. & Soden, B. J. Water vapor feedback and global warming. *Annu. Rev. Energy Environ.* **25,** 441–475 (2000).
5. Sherwood, S. C. & Meyer, C. L. The general circulation and robust relative humidity. *J. Clim.* **19,** 6278–6290 (2006).
6. Robinson, P. J. Temporal trends in United States dew point temperatures. *Int. J. Climatol.* **20,** 985–1002 (2000).
7. Wang, J. X. L. & Gaffen, D. J. Late-twentieth-century climatology and trends of surface humidity and temperature in China. *J. Clim.* **14,** 2833–2845 (2001).
8. Philipona, R., Durr, B., Marty, C., Ohmura, A. & Wild, M. Radiative forcing—measured at Earth's surface—corroborate the increasing greenhouse effect. *Geophys. Res. Lett.* 31, 10.1029/2003GL018765 (2004).
9. Ishii, M., Shouji, A., Sugimoto, S. & Matsumoto, T. Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and Kobe collection. *Int. J. Climatol.* **25,** 865–879 (2005).
10. Dai, A. Recent climatology, variability, and trends in global surface humidity. *J. Clim.* **19,** 3589–3606 (2006).
11. Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* **19,** 5686–5699 (2006).
12. Wentz, F. J., Ricciardulli, L., Hilburn, K. & Mears, C. How much more rain will global warming bring? *Science* **317,** 233–235 (2007).
13. Zhang, X. *et al.* Detection of human influence on twentieth-century precipitation trends. *Nature* **448,** 461–465 (2007).
14. Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419,** 224–232 (2002).
15. Holland, G. J. The maximum potential intensity of tropical cyclones. *J. Atmos. Sci.* **54,** 2519–2541 (1997).
16. Steadman, R. G. A universal scale of apparent temperature. *J. Clim. Appl. Meteorol.* **23,** 1674–1687 (1984).
17. Arnell, N. *et al.* in *Climate Change 2001: Impacts, Adaptation and Vulnerability* (eds McCarthy, J. J. *et al.*) 191–233 (Cambridge Univ. Press, Cambridge, 2001).
18. Gedney, N. *et al.* Detection of a direct carbon dioxide effect in continental river runoff records. *Nature* **439,** 835–838 (2006).
19. Willett, K. M. *Creation and Analysis of HadCRUH: A New Global Surface Humidity Dataset.* PhD thesis, Univ. East Anglia (2007).
20. Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16,** 147–168 (2000).
21. Johns, T. C. *et al.* Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim. Dyn.* **20,** 583–612 (2003).
22. Gillett, N. P., Wehner, M. F., Tett, S. F. B. & Weaver, A. J. Testing the linearity of the response to combined greenhouse gas and sulphate aerosol forcing. *Geophys. Res. Lett.* 31, 10.1029/2004GL020111 (2004).
23. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.* 111, 10.1029/2005JD006548 (2006).
24. Jones, P. D., New, M., Parker, D. E., Martin, S. & Rigor, I. G. Surface air temperature and its changes over the past 150 years. *Rev. Geophys.* **37,** 173–199 (1999).
25. McCarthy, M. P. & Willett, K. M. *Report on Estimates of Observational Uncertainty in Surface Humidity and Free-atmosphere Temperature and Humidity Data.* Technical Report No. 71 (Hadley Centre, Exeter, 2006).
26. Stott, P. A. *et al.* External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290,** 2133–2137 (2000).
27. Allen, M. R. & Stott, P. A. Estimating signal amplitudes in optimal fingerprinting. Part I: Theory. *Clim. Dyn.* **21,** 477–491 (2003).
28. Allen, M. R. & Tett, S. F. B. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15,** 419–434 (1999).
29. Santer, B. D. *et al.* Identification of human-induced changes in atmospheric moisture content. *Proc. Natl Acad. Sci. USA* **104,** 15248–15253 (2007).
30. Trenberth, K. E., Fasullo, J. & Smith, L. Trends and variability in column-integrated atmospheric water vapor. *Clim. Dyn.* **24,** 741–758 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Stott and D. Fereday for help in obtaining HadCM3 control data; B. Santer for advice and discussion; and M. Allen for the use of his optimal detection code. K.M.W. was supported by a CASE studentship from the UK Natural Environment Research Council and the Met Office. N.P.G. and P.D.J. acknowledge support from the Climate Change Detection and Attribution Project, jointly funded by NOAA's Office of Global Programs and the US Department of Energy. N.P.G. also acknowledges the support of the Leverhulme Trust. P.D.J. acknowledges the support of the Office of Science, US Department of Energy. P.W.T. was supported by the Department of the Environment, Food and Rural Affairs.

Author Contributions K.M.W. compiled HadCRUH during a PhD project supervised by the other three authors, and prepared the Methods section. N.P.G. performed the detection and attribution analysis and prepared the remainder of the manuscript. P.D.J. proposed the PhD project, and provided advice and guidance. P.W.T. provided advice and guidance and facilitated access to Met Office observations and model data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.P.G. (n.gillett@uea.ac.uk).

METHODS

HadCRUH is a new data set of monthly mean surface humidity anomalies at $5^\circ \times 5^\circ$ resolution, covering the period 1973–2003, with respect to the 1974–2003 climatology[19] (available at http://www.hadobs.org). It is nearly global in coverage, combining land data from observing stations and marine data from ships, buoys and observing platforms. The land data are sourced from version 2 of the integrated surface data set, supplied by the National Climatic Data Center[31]. The marine data are sourced from release 2.1 of ICOADS (International Comprehensive Ocean-Atmosphere Data Set)[32] for 1973–1997 and from NCEP GTS data (Global Telecommunications System data made available through NOAA's National Centers for Environmental Prediction) for 1998–2003. Spatial coverage and observing frequencies are shown for the land and marine data in Supplementary Fig. 1.

HadCRUH is available in specific humidity (q) and relative humidity (RH). All data originate from simultaneous hourly temperature (T) and dewpoint temperature (T_{dw}) (in °C). Depending on whether the temperature is above or below the freezing point, T_{dw} was first used to calculate vapour pressure with respect to either water (e_w) or ice (e_i)[33]:

$$e_w = 6.1121\left(1.0007 + 3.46 \times 10^{-6} P\right) \exp\left[\frac{(18.729 - T_{dw}/227.3)\,T_{dw}}{257.87 + T_{dw}}\right]$$

$$e_i = 6.1115\left(1.0003 + 4.18 \times 10^{-6} P\right) \exp\left[\frac{(23.036 - T_{dw}/333.7)\,T_{dw}}{279.82 + T_{dw}}\right]$$

The saturation vapour pressure e_s, with repect to water or ice, was calculated by replacing T_{dw} with T in the above equations. q and RH were then found[34]:

$$q = 1000\left[\frac{\varepsilon e}{P - (1-\varepsilon)e}\right]$$

$$\mathrm{RH} = 100e/e_s$$

where $e = e_w$ if $T > 0$ °C and $e = e_i$ if $T < 0$ °C; $\varepsilon = 0.622$; and $P = 1{,}013 - Z_s/10$, where Z_s is the height in metres at which the observation was made.

All land data were quality controlled, taking into account issues specific to humidity such as drying and freezing of the wet-bulb thermometer, and converted into pentad mean (five-day average) q and RH station time series. The quality control procedure removed about 10% of all values. The land data were subsequently homogenized by identifying and adjusting breakpoints relative to a background composite neighbour series. The process of breakpoint identification was based on a Kolomogorov–Smirnov test applied to the station minus neighbour composite series and a subjective assessment of the break's origin. Adjustments were made relative to the composite neighbour series rather than to the station series in isolation, to retain any climate change signal better. This method was shown to be effective at removing spurious jumps without introducing additional jumps for a set of case study stations. The land data were gridded by using simple (non-weighted) averaging over each grid box. The resulting gridded products were more spatially homogeneous in both q and RH than those derived from the raw data.

The marine data were checked for homogeneity, and observations were removed that lay outside a plausible range compared with neighbouring observations, following the convention of the Hadley Centre Marine Data System[35] (MDS). Grid box means for the marine data were created in accordance with the MDS, with the effects of outliers reduced by 'winsorizing'[35]. For grid boxes containing both land and ocean, each component value was weighted according to the proportion of land in the grid box. However, neither component was allowed to be weighted less than 25% (ref. 36).

31. Lott, N., Baldwin, R. & Jones, P. *The FCC Integrated Surface Hourly Database, a New Resource of Global Climate Data.* Technical Report No. 2001–01 (National Climatic Data Center, Asheville, NC, 2001).
32. Worley, S. J., Woodruff, S. D., Reynolds, R. W., Lubker, S. J. & Lott, N. ICOADS release 2.1 data and products. *Int. J. Climatol.* **25,** 823–842 (2005).
33. Buck, A. L. New equations for computing vapor-pressure and enhancement factor. *J. Appl. Met.* **20,** 1527–1532 (1981).
34. Peixoto, J. P. & Oort, A. H. The climatology of relative humidity in the atmosphere. *J. Clim.* **9,** 3443–3463 (1996).
35. Rayner, N. A. *et al.* Improved analyses of changes and uncertainties in sea surface temperature measured *in situ* since the mid-nineteenth century: The HadSST2 dataset. *J. Clim.* **19,** 446–469 (2006).
36. Jones, P. D. *et al.* Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.* **106,** 3371–3380 (2001).

LETTERS

Quantifying the evolutionary dynamics of language

Erez Lieberman[1,2,3]*, Jean-Baptiste Michel[1,4]*, Joe Jackson[1], Tina Tang[1] & Martin A. Nowak[1]

Human language is based on grammatical rules[1–4]. Cultural evolution allows these rules to change over time[5]. Rules compete with each other: as new rules rise to prominence, old ones die away. To quantify the dynamics of language evolution, we studied the regularization of English verbs over the past 1,200 years. Although an elaborate system of productive conjugations existed in English's proto-Germanic ancestor, Modern English uses the dental suffix, '-ed', to signify past tense[6]. Here we describe the emergence of this linguistic rule amidst the evolutionary decay of its exceptions, known to us as irregular verbs. We have generated a data set of verbs whose conjugations have been evolving for more than a millennium, tracking inflectional changes to 177 Old-English irregular verbs. Of these irregular verbs, 145 remained irregular in Middle English and 98 are still irregular today. We study how the rate of regularization depends on the frequency of word usage. The half-life of an irregular verb scales as the square root of its usage frequency: a verb that is 100 times less frequent regularizes 10 times as fast. Our study provides a quantitative analysis of the regularization process by which ancestral forms gradually yield to an emerging linguistic rule.

Natural languages comprise elaborate systems of rules that enable one speaker to communicate with another[7]. These rules serve to simplify the production of language and enable an infinite array of comprehensible formulations[8–10]. However, each rule has exceptions, and even the rules themselves wax and wane over centuries and millennia[11,12]. Verbs that obey standard rules of conjugation in their native language are called regular verbs[13]. In the Modern English language, regular verbs are conjugated into the simple past and past-participle forms by appending the dental suffix '-ed' to the root (for instance, infinitive/simple past/past participle: talk/talked/talked). Irregular verbs obey antiquated rules (sing/sang/sung) or, in some cases, no rule at all (go/went)[14,15]. New verbs entering English universally obey the regular conjugation (google/googled/googled), and many irregular verbs eventually regularize. It is much rarer for regular verbs to become irregular: for every 'sneak' that 'snuck' in[16], there are many more 'flews' that 'flied' out.

Although less than 3% of modern verbs are irregular, the ten most common verbs are all irregular (be, have, do, go, say, can, will, see, take, get). The irregular verbs are heavily biased towards high frequencies of occurrence[17,18]. Linguists have suggested an evolutionary hypothesis underlying the frequency distribution of irregular verbs: uncommon irregular verbs tend to disappear more rapidly because they are less readily learned and more rapidly forgotten[19,20].

To study this phenomenon quantitatively, we studied verb inflection beginning with Old English (the language of *Beowulf*, spoken around AD 800), continuing through Middle English (the language of Chaucer's *Canterbury Tales*, spoken around AD 1200), and ending with Modern English, the language as it is spoken today. The modern '-ed' rule descends from Old English 'weak' conjugation, which applied to three-quarters of all Old English verbs[21]. The exceptions—ancestors of the modern irregular verbs—were mostly members of the so-called 'strong' verbs. There are seven different classes of strong verbs with exemplars among the Modern English irregular verbs, each with distinguishing markers that often include characteristic vowel shifts. Although stable coexistence of multiple rules is one possible outcome of rule dynamics, this is not what occurred in English verb inflection[22]. We therefore define regularity with respect to the modern '-ed' rule, and call all these exceptional forms 'irregular'.

We consulted a large collection of grammar textbooks describing verb inflection in these earlier epochs, and hand-annotated every irregular verb they described (see Supplementary Information). This provided us with a list of irregular verbs from ancestral forms of English. By eliminating verbs that were no longer part of Modern English, we compiled a list of 177 Old English irregular verbs that remain part of the language to this day. Of these 177 Old English irregulars, 145 remained irregular in Middle English and 98 are still irregular in Modern English. Verbs such as 'help', 'grip' and 'laugh', which were once irregular, have become regular with the passing of time.

Next we obtained frequency data for all verbs by using the CELEX corpus, which contains 17.9 million words from a wide variety of textual sources[23]. For each of our 177 verbs, we calculated the frequency of occurrence among all verbs. We subdivided the frequency spectrum into six logarithmically spaced bins from 10^{-6} to 1. Figure 1a shows the number of irregular verbs in each frequency bin. There are only two verbs, 'be' and 'have', in the highest frequency bin, with mean frequency $>10^{-1}$. Both remain irregular to the present day. There are 11 irregular verbs in the second bin, with mean frequency between 10^{-2} and 10^{-1}. These 11 verbs have all remained irregular from Old English to Modern English. In the third bin, with a mean frequency between 10^{-3} and 10^{-2}, we find that 37 irregular verbs of Old English all remained irregular in Middle English, but only 33 of them are irregular in Modern English. Four verbs in this frequency range, 'help', 'reach', 'walk' and 'work', underwent regularization. In the fourth frequency bin, 10^{-4} to 10^{-3}, 65 irregular verbs of Old English have left 57 in Middle and 37 in Modern English. In the fifth frequency bin, 10^{-5} to 10^{-4}, 50 irregulars of Old English have left 29 in Middle and 14 in Modern English. In the sixth frequency bin, 10^{-6} to 10^{-5}, 12 irregulars of Old English decline to 9 in Middle and only 1 in Modern English: 'slink', a verb that aptly describes this quiet process of disappearance (Table 1).

Plotting the number of irregular verbs against their frequency generates a unimodal distribution with a peak between 10^{-4} and 10^{-3}. This unimodal distribution again demonstrates that irregular verbs are not an arbitrary subset of all verbs, because a random subset of verbs (such as all verbs that contain the letter 'm') would follow Zipf's law, a power law with a slope of -0.75 (refs 24,25).

[1]Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, [2]Department of Applied Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work

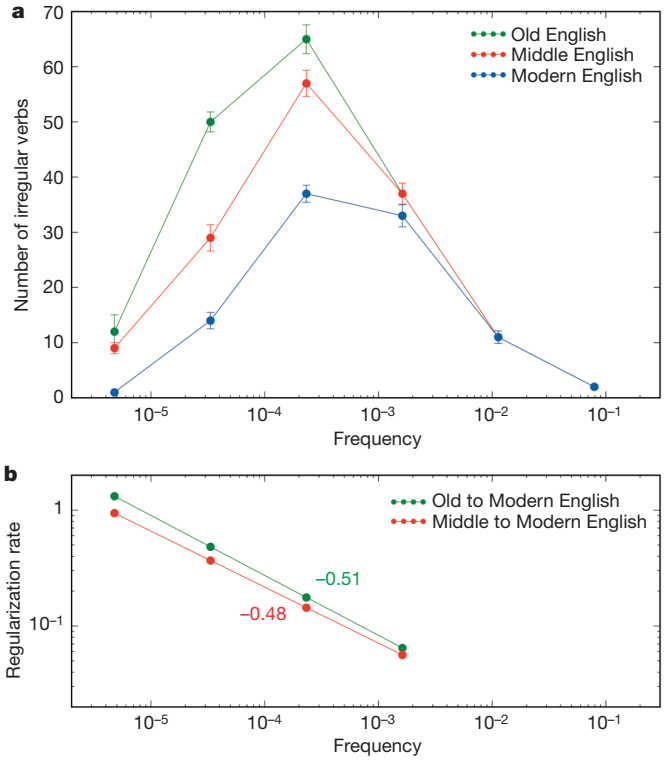


Figure 1 | Irregular verbs regularize at a rate that is inversely proportional to the square root of their usage frequency. **a**, The evolution of 177 verbs from Old English (green) over time, through Middle English (red) and Modern English (blue). The fraction remaining irregular in each bin decreases as the frequency decreases. The frequency shown is that of the modern descendant, and was computed using the CELEX corpus. Error bars indicate standard deviation and were calculated using the bootstrap method. **b**, The regularization rate of irregular verbs as a function of frequency. The relative regularization rates obtained by comparing Old versus Modern English (green) and Middle versus Modern English (red) scale linearly on a log–log plot with a downward slope of nearly one-half. The regularization rate and the half-life scale with the square root of the frequency.

Four of our six frequency bins, those between 10^{-6} and 10^{-2}, allow us to estimate the relative regularization rates of irregular verbs. Calculating the relative regularization rates of verbs of different frequencies is independent of time, which makes the dating of Old and Middle English irrelevant for this calculation. We can plot regularization rate versus frequency and fit a straight line in a log–log plot (Fig. 1b). By comparing Old and Modern English we obtain a slope of about -0.51. Therefore, an irregular verb that is 100 times less frequent is regularized 10 times as fast. In other words, the half-life of irregular verbs is proportional to the square root of their frequency. By comparing Middle and Modern English we find a slope of about -0.48, consistent with the previous result. Both comparisons show that low-frequency irregular verbs are selectively forgotten.

Figure 2a shows the exponential decay of the irregular verbs in the four frequency bins between 10^{-6} and 10^{-2} as a function of time. From these data, which depend on the dating of Old and Middle English, we can estimate actual half-lives of the irregular verbs in different frequency bins. Irregular verbs that occur with a frequency between 10^{-6} and 10^{-5} have a half-life of about 300 years, whereas those with a frequency between 10^{-4} and 10^{-3} have a half-life of 2,000 years. If we fit half-life versus frequency with a straight line in a log–log plot, we obtain a slope of 0.50, which again suggests that the half-life of irregular verbs is proportional to approximately the square root of their frequency (Fig. 2b). It is noteworthy that various methods of fitting the data give the same results.

We cannot directly determine the regularization rate for frequency bins above 10^{-2}, because regularization is so slow that no event was observed in the time span of our data; however, we can extrapolate. For instance, the half-life of verbs with frequencies between 10^{-2} and 10^{-1} should be 14,400 years. For these bins, the population is so small and the half-life so long that we may not see a regularization event in the lifetime of the English language.

To test whether the dynamics within individual competing rules were captured by our global analysis, we studied the decay of individual classes of strong verbs (for example, hit/hit/hit, hurt/hurt/hurt; draw/drew/drawn, grow/grew/grown)[26]. Although our resolution is limited by the small sample size, exponential decay is once again observed, with similar exponents (see Supplementary Fig. 1). Like a Cheshire cat, dying rules vanish one instance at a time, leaving behind a unimodal frown.

Because adequate corpora of Old and Middle English do not exist, we have estimated the frequency of an irregular verb of Old and Middle English by the frequency of the corresponding (regular or irregular) verb of Modern English[27]. A large fraction of verbs would have had to change frequency by several orders of magnitude to

Table 1 | The 177 irregular verbs studied

Frequency	Verbs	Regularization (%)	Half-life (yr)
10^{-1}–1	be, have	0	38,800
10^{-2}–10^{-1}	come, do, find, get, give, go, know, say, see, take, think	0	14,400
10^{-3}–10^{-2}	begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, help, hold, leave, let, lie, lose, reach, rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, walk, win, work, write	10	5,400
10^{-4}–10^{-3}	arise, bake, bear, beat, bind, bite, blow, bow, burn, burst, carve, chew, climb, cling, creep, dare, dig, drag, flee, float, flow, fly, fold, freeze, grind, leap, lend, lock, melt, reckon, ride, rush, shape, shine, shoot, shrink, sigh, sing, sink, slide, slip, smoke, spin, spring, starve, steal, step, stretch, strike, stroke, suck, swallow, swear, sweep, swim, swing, tear, wake, wash, weave, weep, weigh, wind, yell, yield	43	2,000
10^{-5}–10^{-4}	bark, bellow, bid, blend, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave, knead, low, milk, mourn, mow, prescribe, redden, reek, row, scrape, seethe, shear, shed, shove, slay, slit, smite, sow, span, spurn, sting, stink, strew, stride, swell, tread, uproot, wade, warp, wax, wield, wring, writhe	72	700
10^{-6}–10^{-5}	bide, chide, delve, flay, hew, rue, shrive, slink, snip, spew, sup, wreak	91	300

177 Old English irregular verbs were compiled for this study. These are arranged according to frequency bin, and in alphabetical order within each bin. Also shown is the percentage of verbs in each bin that have regularized. The half-life is shown in years. Verbs that have regularized are indicated in red. As we move down the list, an increasingly large fraction of the verbs are red; the frequency-dependent regularization of irregular verbs becomes immediately apparent.

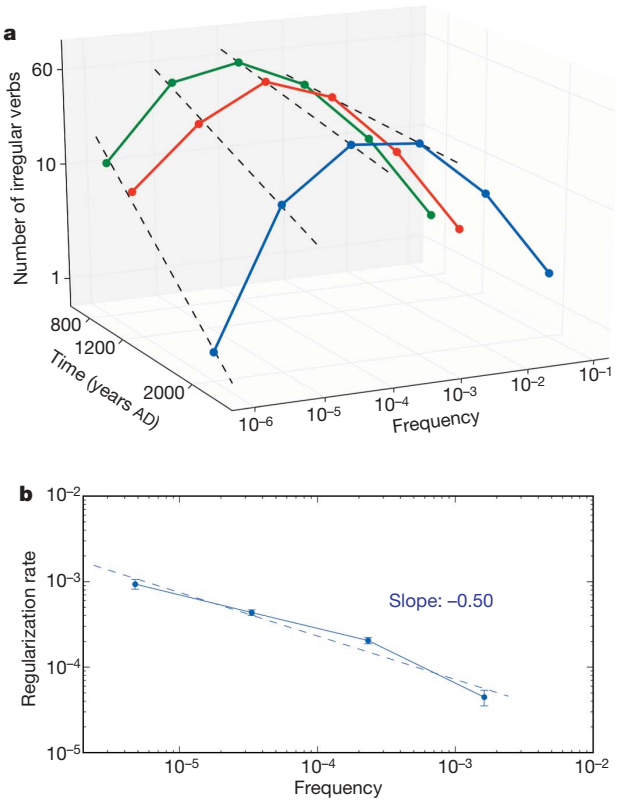


Figure 2 | Irregular verbs decay exponentially over time. **a**, Specifying approximate dates of Old and Middle English allows computation of absolute regularization rates. Regularization rates increase as frequencies decrease, but are otherwise constant over time. **b**, Absolute rates of regularization are shown as a function of frequency. Error bars indicate standard deviation and were calculated using the bootstrap method. The square-root scaling is obtained again.

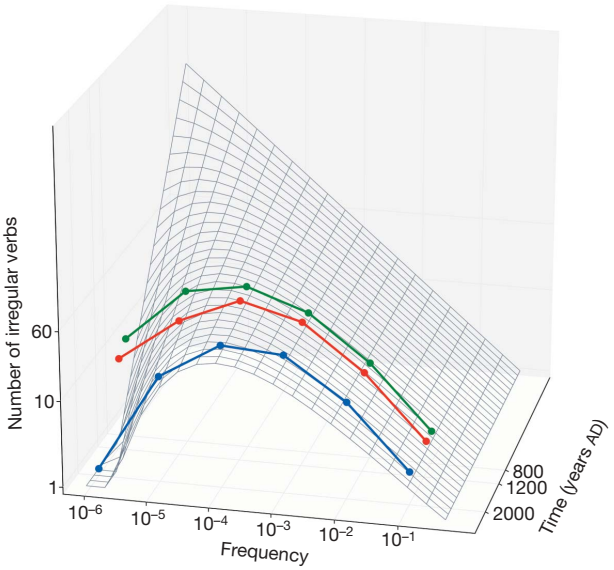


Figure 3 | Extrapolating forward and backward in time using the observation that regularization rate scales as the square root of frequency. The differential system is exactly solvable and the solution fits all three observed distributions. As we move backward in time, the distribution of irregular verbs approaches the zipfian distribution characteristic of random sets of words. The distribution for exceptions to the '-ed' rule became non-random because of frequency-dependent regularization due to selective pressure from the emerging rule.

interfere with the effects observed. To verify that large changes in frequency are rare, we compared frequency data from CELEX with frequencies drawn from the largest available corpus of Middle English texts[28]. Of 50 verbs, only 5 had frequency changes greater than a factor of 10 (Supplementary Fig. 2).

Our analysis covers a vast period, spanning the Norman invasion and the invention of the printing press, but these events did not upset the dynamics of English regularization. Therefore, it is possible to retrospectively trace the evolution of the irregular verbs, moving backwards in time from the observed Modern English distribution and up through Middle and Old English. Going still further back in time allows us to explore the effects of completely undoing the frequency-dependent selective process that the irregular verbs have undergone. Eventually, the shape of the curve changes from unimodal to a power law decline, with slope of nearly −0.75 (Fig. 3). This finding is consistent with the fact that random subsets of verbs (and of all types of words) exhibit such a zipfian distribution. The observed irregular verb distribution is the result of selective pressure on a random collection of ancestral verbs.

We can also make predictions about the future of the past tense. By the time one verb from the set 'begin, break, bring, buy, choose, draw, drink, drive, eat, fall' will regularize, five verbs from the set 'bid, dive, heave, shear, shed, slay, slit, sow, sting, stink' will be regularized. If the current trends continue, only 83 of the 177 verbs studied will be irregular in 2500.

What will be the next irregular verb to regularize? It is likely to be wed/wed/wed. The frequency of 'wed' is only 4.2 uses per million verbs, ranking at the very bottom of the modern irregular verbs. Indeed, it is already being replaced in many contexts by wed/wedded/wedded. Now is your last chance to be a 'newly wed'. The married couples of the future can only hope for 'wedded' bliss.

In previous millennia, many rules vied for control of English language conjugation, and fossils of those rules remain to this day. Yet, from this primordial soup of conjugations, the dental suffix '-ed' emerged triumphant. The competing rules are long dead, and unfamiliar even to well-educated native speakers. These rules disappeared because of the gradual erosion of their instances by a process that we call regularization. But regularity is not the default state of a language—a rule is the tombstone of a thousand exceptions.

METHODS SUMMARY

We searched 11 reference works on Old and Middle English, compiling a list of every irregular verb that we found. We determined whether each verb is still present in Modern English. For all Old English verbs whose descendants remained in the English language, we checked whether they were still irregular using a complete listing of the modern irregular verbs. If they had regularized, we determined when regularization had occurred on the basis of the last time period in which we found a positive annotation listing the verb as irregular. A list of sources used and the entire resulting annotation are provided in the Supplementary Information.

We determined usage frequencies for all the verbs using the CELEX database. We then binned the Old English irregular verbs using a standard logarithmic binning algorithm in Python. We used the resulting binning to determine regularization rates for verbs of differing frequencies. Regularization rates (Fig. 1b) for each bin were computed directly. The fits to exponential decay (Fig. 2) and to the solution of the irregular equation (Fig. 3 and Supplementary Information) were produced using the method of least squares. The Python source code for producing the figures and the table is available at http://www.languagedata.org.

Received 20 March; accepted 27 July 2007.

1. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, Cambridge, 1965).
2. Lightfoot, D. *The Development of Language: Acquisition, Change and Evolution* (Blackwell, Oxford, 1999).
3. Clark, R. & Roberts, I. A computational model of language learnability and language change. *Linguist. Inq.* **24,** 299–345 (1993).
4. Abrams, D. & Strogatz, S. Modelling the dynamics of language death. *Nature* **424,** 900 (2003).
5. Nowak, M. A., Komarova, N. L. & Niyogi, P. Computational and evolutionary aspects of language. *Nature* **417,** 611–617 (2002).

6. Hooper, J. in *Current Progress in Historical Linguistics* (ed. Christie, W.) 95–105 (North-Holland, Amsterdam, 1976).
7. Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: what is it, who has it, and how did it evolve? *Science* **298,** 1569–1579 (2002).
8. Chomsky, N. & Lasnik, H. in *Syntax: An International Handbook of Contemporary Research* (ed. Jacobs, J.) 506–569 (de Gruyte, Berlin, 1993).
9. Dougherty, R. C. *Natural Language Computing* (Lawrence Erlbaum, Hillsdale, 1994).
10. Stabler, E. P. & Keenan, E. L. Structural similarity within and among languages. *Theor. Comput. Sci.* **293,** 345–363 (2003).
11. Niyogi, P. *The Computational Nature of Language Learning and Evolution* (MIT Press, Cambridge, 2006).
12. Labov, W. Transmission and diffusion. *Language* **83,** 344–387 (2007).
13. Pinker, S. *Words and Rules: The Ingredients of Language* (Basic Books, New York, 1999).
14. Kroch, A. Reflexes of grammar in patterns of language change. *Lang. Var. Change* **1,** 199–244 (1989).
15. Kroch, A. in *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory* (eds Beals, K. *et al.*) 180–201 (CLS, Chicago, 1994).
16. Pinker, S. The irregular verbs. *Landfall* 83–85 (Autumn issue, 2000).
17. Bybee, J. *Morphology: a Study of Relation Between Meaning and Form* (Benjamins, Amsterdam, 1985).
18. Greenberg, J. in *Current Trends in Linguistics III* (eds Sebeok, T. A. *et al.*) 61–112 (Mouton, The Hague, 1966).
19. Bybee, J. From usage to grammar: the mind's response to repetition. *Language* **82,** 711–733 (2006).
20. Corbett, G., Hippisley, A., Brown, D. & Marriott, P. in *Frequency and the Emergence of Linguistic Structure* (eds Bybee, J. & Hopper, P.) 201–226 (Benjamins, Amsterdam, 2001).
21. Hare, M. & Elman, J. Learning and morphological change. *Cognition* **56,** 61–98 (1995).
22. Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R. & Pinker, S. German inflection: the exception that proves the rule. *Cognit. Psychol.* **29,** 189–256 (1995).
23. Van der Wouden, T. in *Papers from the 3rd International EURALEX Congress* (eds Magay, T. & Zigány, J.) 363–373 (Akadémiai Kiadó, Budapest, 1988).
24. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, 1949).
25. Miller, G. A. Some effects of intermittent silence. *Am. J. Psychol.* **70,** 311–314 (1957).
26. Yang, C. *Knowledge and Learning in Natural Language* (Oxford Univ. Press, New York, 2002).
27. Glushko, M. Towards the quantitative approach to studying evolution of English verb paradigm. *Proc. 19th Scand. Conf. Ling.* **31,** 30–45 (2003).
28. Kroch, A. & Taylor, A. *Penn-Helsinki Parsed Corpus of Middle English* [CD-ROM] 2nd edn (2000) ⟨http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-2/⟩.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the John Templeton Foundation and by a grant from the NSF-NIH joint programme in mathematical biology. The Program for Evolutionary Dynamics is sponsored by J. Epstein. E.L. was supported by the National Defense Science and Engineering Graduate Fellowship and the National Science Foundation Graduate Fellowship. We thank S. Pinker, J. Rau, D. Donoghue and A. Presser for discussions, and J. Saragosti for help with visualization.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.L. (erez@erez.com).

LETTERS

Frequency of word-use predicts rates of lexical evolution throughout Indo-European history

Mark Pagel[1,2], Quentin D. Atkinson[1] & Andrew Meade[1]

Greek speakers say "*ουρά*", Germans "*schwanz*" and the French "*queue*" to describe what English speakers call a 'tail', but all of these languages use a related form of 'two' to describe the number after one. Among more than 100 Indo-European languages and dialects, the words for some meanings (such as 'tail') evolve rapidly, being expressed across languages by dozens of unrelated words, while others evolve much more slowly—such as the number 'two', for which all Indo-European language speakers use the same related word-form[1]. No general linguistic mechanism has been advanced to explain this striking variation in rates of lexical replacement among meanings. Here we use four large and divergent language corpora (English[2], Spanish[3], Russian[4] and Greek[5]) and a comparative database of 200 fundamental vocabulary meanings in 87 Indo-European languages[6] to show that the frequency with which these words are used in modern language predicts their rate of replacement over thousands of years of Indo-European language evolution. Across all 200 meanings, frequently used words evolve at slower rates and infrequently used words evolve more rapidly. This relationship holds separately and identically across parts of speech for each of the four language corpora, and accounts for approximately 50% of the variation in historical rates of lexical replacement. We propose that the frequency with which specific words are used in everyday language exerts a general and law-like influence on their rates of evolution. Our findings are consistent with social models of word change that emphasize the role of selection, and suggest that owing to the ways that humans use language, some words will evolve slowly and others rapidly across all languages.

Languages, like species, evolve by way of a process of descent with modification (Supplementary Table 1). The remarkable diversity of languages—there are about 7,000 known living languages[7]—is a product of this process acting over thousands of years. Ancestral languages split to form daughter languages that slowly diverge as shared lexical, phonological and grammatical features are replaced by novel forms. In the study of lexical change, the basic unit of analysis is the cognate. Cognates are words of similar meaning with systematic sound correspondences indicating they are related by common ancestry. For example, cognates meaning 'water' exist in English (water), German (*wasser*), Swedish (*vatten*) and Gothic (*wato*), reflecting descent from proto-Germanic (**water*).

Early lexicostatistical[8] studies of Malayo-Polynesian and Indo-European language families revealed that the rate at which new cognates arise varies across meaning categories[1,9]. More recently we have obtained direct estimates of rates of cognate replacement on linguistic phylogenies (family trees) of Indo-European and Bantu languages, using a statistical model of word evolution in a bayesian Markov chain Monte Carlo (MCMC) framework[10]. We found that rates of cognate replacement varied among meanings, and that rates for different meanings in Indo-European were correlated with their paired meanings in the Bantu languages. This indicates that variation in the rates of lexical replacement among meanings is not merely an historical accident, but rather is linked to some general process of language evolution.

Social and demographic factors proposed to affect rates of language change within populations of speakers include social status[11], the strength of social ties[12], the size of the population[13] and levels of outside contact[14]. These forces may influence rates of evolution on a local and temporally specific scale, but they do not make general predictions across language families about differences in the rate of lexical replacement among meanings. Drawing on concepts from theories of molecular[15] and cultural evolution[16–18], we suggest that the frequency with which different meanings are used in everyday language may affect the rate at which new words arise and become adopted in populations of speakers. If frequency of meaning-use is a shared and stable feature of human languages, then this could provide a general mechanism to explain the large differences across meanings in observed rates of lexical replacement. Here we test this idea by examining the relationship between the rates at which Indo-European language speakers adopt new words for a given meaning and the frequency with which those meanings are used in everyday language.

We estimated the rates of lexical evolution for 200 fundamental vocabulary meanings[8] in 87 Indo-European languages[6]. Rates were estimated using a statistical likelihood model of word evolution[10] applied to phylogenetic trees of the 87 languages (Supplementary Fig. 1). The number of cognates observed per meaning varied from one to forty-six. For each of the 200 meanings, we calculated the mean of the posterior distribution of rates as derived from a bayesian MCMC model that simultaneously accounts for uncertainty in the parameters of the model of cognate replacement and in the phylogenetic tree of the languages (Methods). Rate estimates were scaled to represent the expected number of cognate replacements per 10,000 years, assuming a 8,700-year age for the Indo-European language family[6]. Opinions on the age of Indo-European vary between approximately 6,000 and 10,000 years before present[19,20]. Using a different calibration would change the absolute values of the rates but not their relative values.

Figure 1a shows the inferred distribution of rate estimates, where we observe a roughly 100-fold variation in rates of lexical evolution among the meanings. At the slow end of the distribution, the rates predict zero to one cognate replacements per 10,000 years for words such as 'two', 'who', 'tongue', 'night', 'one' and 'to die'. By comparison, for the faster evolving words such as 'dirty', 'to turn', 'to stab' and 'guts', we predict up to nine cognate replacements in the same time period. In the historical context of the Indo-European language family, this range yields an expectation of between 0–1 and 43 lexical replacements throughout the ~130,000 language-years of evolution the linguistic tree represents, very close to the observed range in the

[1]School of Biological Sciences, University of Reading, Whiteknights, Reading, Berkshire, RG6 6AS, UK. [2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.

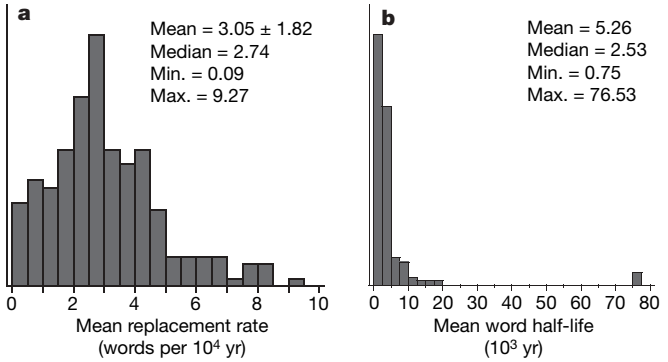


Figure 1 | Frequency plots for rates of lexical evolution in Indo-European across 200 fundamental vocabulary meanings. **a**, The mean estimated rate of cognate replacement for each meaning. **b**, The same rate distribution converted to word half-lives[10], or the time in which there is a 50% chance the word will be replaced by a different non-cognate form. The longest half-lives (76,530 years) are for meanings that show no change across Indo-European (Supplementary Information).

fundamental vocabulary of 1–46 distinct cognate classes among the different meanings. These rates can be converted to estimates of the linguistic half-life[10] (Methods), or the time in which there is a 50% chance the word will be replaced by a different non-cognate form. These times vary from 750 years for the fastest evolving words to over 10,000 years for the slowest (Fig. 1b).

We used spoken and written language corpus data from English[2], Spanish[3], Russian[4] and Greek[5] to measure the frequency of meaning-use (Supplementary Table 2). These languages sample from across the Indo-European language family (Supplementary Fig. 1), and their corpora were selected to provide large samples of language use (20–100 million words each). Figure 2 shows that the distribution of word-use frequencies in each language is highly skewed, such that most words are used relatively infrequently (fewer than 100 times per million words), with a small number of frequently used words (as often as 35,000 times per million words) accounting for most speech. Word-use frequencies are highly correlated among the four languages ($0.78 < r < 0.89$, mean $r = 0.84$; Supplementary Fig. 2), showing that words used at a high frequency in one language tend to be used at a high frequency in the other languages. Because the four languages span the Indo-European tree, this suggests that frequency values are representative of Indo-European language use, and that frequencies of meaning-use have been remarkably stable throughout Indo-European history (Supplementary Fig. 1).

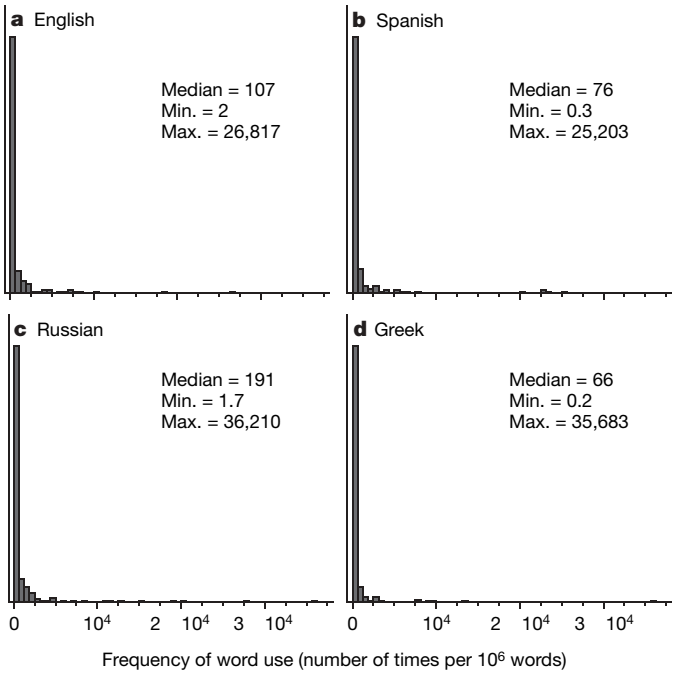


Figure 2 | Distribution of frequency of meaning-use for 200 meanings in four Indo-European languages. **a**, English; **b**, Spanish; **c**, Russian; and **d**, Greek. These four languages sample from across the Indo-European language family. Supplementary Fig. 1 shows where each language fits on the Indo-European phylogeny. Word-use frequencies are highly correlated among the four languages ($0.78 < r < 0.89$, mean $r = 0.84$; Supplementary Fig. 2).

Figure 3 plots the rate of lexical replacement against frequency of word-use for the 200 meanings. Separately in each corpus we observe a negative relationship (bold black line) between word frequency and rate (English, $r = -0.37$; Spanish, $r = -0.35$; Russian, $r = -0.41$; and Greek, $r = -0.32$: all $P < 0.0001$). In all four languages, the more a meaning is used today, the slower its rate of evolution has been throughout the 6,000- to 10,000-year history of Indo-European.

Some parts of speech are used more than others, so it is possible that the observed relationship arises from an effect of part of speech on rates of evolution. To examine this effect, we categorized meanings as either nouns, adjectives, verbs, pronouns, numbers, conjunctions, prepositions or special adverbs ('what', 'when', 'where', 'how', 'here', 'there' and 'not'). We then predicted variation in rates of lexical replacement from a regression model allowing both of these effects to operate simultaneously, thereby controlling for one another. The inverse relationship between frequency of meaning-use and rate of lexical evolution holds separately for parts of speech

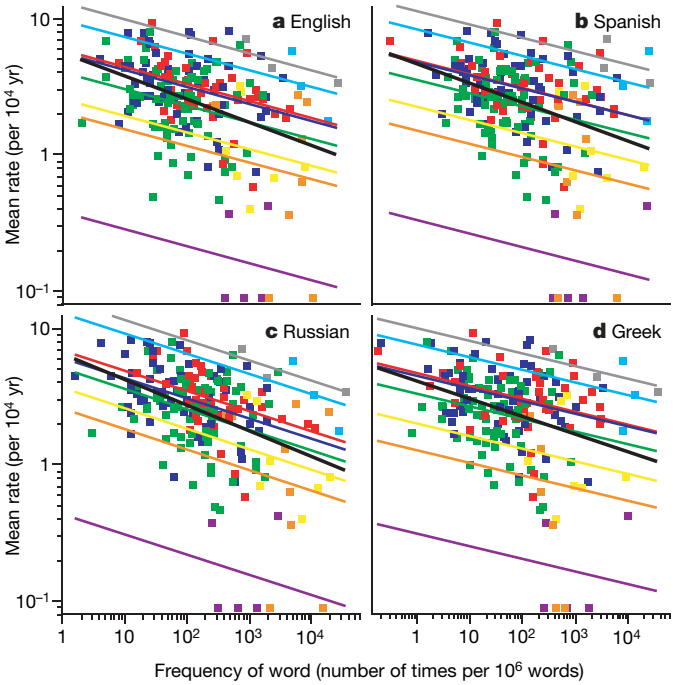


Figure 3 | Frequency of meaning-use plotted against estimated rate of lexical evolution for 200 basic meanings in four Indo-European languages. **a**, English; **b**, Spanish; **c**, Russian; and **d**, Greek. A linear regression (bold black line) reveals a consistent negative relationship between log(frequency of meaning-use) and log(rate of lexical replacement) across all four languages (English, $r = 0.37$; Spanish, $r = 0.35$; Russian, $r = 0.41$; and Greek, $r = 0.32$). Points are colour coded according to part of speech (see below). Coloured lines show the results of a multiple regression including frequency and part of speech. All relationships are negative (English, $R = 0.69$; Spanish, $R = 0.69$; Russian, $R = 0.71$; and Greek, $R = 0.69$). The height of each line above the x-axis indicates the relative speed of lexical evolution for a given frequency of meaning-use for each part of speech. Conjunctions (grey) evolve fastest, followed by prepositions (turquoise), adjectives (red), verbs (blue), nouns (green), special adverbs (yellow), pronouns (orange) and numbers (purple).

(Fig. 3, coloured lines). For a given frequency of meaning-use, prepositions and conjunctions evolve most quickly, followed by progressively slower evolution for adjectives, verbs, nouns, special adverbs, pronouns and finally numbers. The rank order of effects for part of speech is identical across the four corpora, and the combined models account for approximately 50% of the variance in rates of lexical evolution (English, $R = 0.69$; Spanish, $R = 0.69$; Russian, $R = 0.71$; and Greek, $R = 0.69$: all $P < 0.0001$; R denotes correlation derived from multiple regression). Adding an interaction effect between part of speech and frequency of meaning-use did not improve the fit of the model: frequency of meaning-use affects the rate of evolution in the same way for each part of speech.

The consistent pattern across the four languages in the relative rates of lexical replacement among parts of speech may help us to understand the mechanism by which word-use frequency affects rate of evolution. Frequency of word-use could directly modify the rate at which new word forms arise, with fewer spontaneous errors occurring for highly expressed words. Errors in word perception, recall and production have been shown to decrease with word frequency according to the 'power law of learning'[21,22]. Alternatively, the rate at which new forms appear could be the same for all meanings, with frequency of use affecting the probability that a population of speakers will come to adopt a given innovation. This suggests that some form of linguistic, frequency-dependent, purifying selection is responsible for the slow rate of evolution of highly expressed words. Innovations, being rare, may not be favoured in speech because there is an increased chance that they will be misinterpreted. The effect should be stronger the more often the meaning is required in speech or the more important it is to the meaning of speech.

In each of our language corpora, numbers, pronouns and the special adverbs evolve the most slowly for a given frequency of word-use. These parts of speech seem important to the meaning of spoken communication, and may therefore be subject to stronger selection. The rapidly evolving parts of speech include conjunctions, prepositions and adjectives whose exact forms may often be less important to conveying meaning. There is also evidence in natural populations of speakers that when more than one word is used to express the same meaning, the relative frequencies of use of the rare words is lower than expected from a neutral drift model of evolution[23], consistent with selection against innovations. These findings may then indicate that the purifying selection model of word evolution provides a more accurate description of how words evolve within populations of speakers. They are also consistent with models of cultural and linguistic evolution that incorporate a conformist bias[16,17], although these models cannot identify a priori which words will be subjected most strongly to such effects.

Our findings, based on a sample of fundamental vocabulary items, identify a general mechanism of linguistic evolution, which is expected to operate across all languages and timescales and makes predictions about rates associated with specific meanings. To the extent that the structure and everyday functions of human verbal communication mean that some words will tend to be used more frequently in all languages, we expect these words to evolve slowly, and vice versa for infrequently used words. Combined with parts of speech, this simple factor allows us to account for about 50% of the variance in rates of lexical replacement throughout the 6,000- to 10,000-year history of Indo-European languages. Given the many social, cultural and cognitive factors that can influence language[11–14], it is striking that word-use frequency alone can explain such a large proportion of the historical variation in rates of evolution. The generality of this influence is suggested in the finding that estimates of the rate of lexical replacement in Indo-European languages are correlated with rate estimates in Bantu[10], Cushitic and Malayo-Polynesian[1].

Being able to link variation in rates of lexical replacement to the frequency of word-use also provides insights into some features of comparative linguistics. One is that we expect languages to diverge initially in the least frequently used parts of their vocabularies. This may mean that languages retain mutual intelligibility far longer than expected from simple uniform rates models of linguistic divergence[8]. Within English, for example, words spoken at a higher frequency are more likely to be of Old English origin[24]. Related to this, the words for frequently used meanings should, on average, be less prone to borrowing during language contact. Higher frequency words may also be more likely to exhibit ancestral morphology. Irregular verbs in English often retain their ancestral morphology, and are among the most frequently expressed verbs[25]. Finally, we note that our rate estimates show that some words evolve slowly enough to allow homologous lexical forms to persist for tens of thousands of years. These slow rates demonstrate that humans are capable of producing a culturally transmitted replicator that, perhaps because of the purifying force of spoken word frequency, can have a replication accuracy as high as that of some genes[26]. Along with continued efforts at identifying cognate words separated by thousands of years of sound change[27], this raises the possibility of using selected lexical items to evaluate hypothesized 'long-range' linguistic relationships such as Eurasiatic[28] and Nostratic[29].

METHODS SUMMARY

Cognate data. We grouped the words for each of the 200 meanings in the fundamental vocabulary of ref. 8, based on previously published Indo-European lexical data[6] (Methods). Meanings had between 1 and 46 cognate sets across the 87 languages in our study, producing a total of 4,049 cognates (including unique word-forms).

Phylogenetic trees. We inferred a bayesian posterior distribution of phylogenetic trees of the 87 languages from a binary data matrix derived from the cognacy classifications[30]. The 4,049 binary vectors in the matrix code for the presence ('1') or absence ('0') of each of the 4,049 cognates. The consensus tree of this posterior sample is reported in Supplementary Fig. 1.

Rates of lexical replacement. We categorized words for each of the 200 meanings into k states representing the k different cognate classes identified for that meaning. For example, $k = 24$ for the meaning class 'big' because it has 24 cognates among the 87 languages. For each meaning, we estimate the instantaneous transition rate, q, from any state (cognate class) i to any state j, as the mean of its bayesian posterior distribution of rates, summed over models of evolution and phylogenetic trees in the posterior sample of trees. This accounts for uncertainty in the model of evolution and in the phylogenetic tree, and does not suffer from loss of information owing to the conversion of cognate data to pair-wise similarity scores between languages. Half-life estimates were derived as described in ref. 10.

Word frequency data. We obtained word-use frequencies from English[2], Spanish[3], Russian[4] and Greek[5] corpora, combining the frequencies for all words comprising a shared canonical form (for example, 'push', 'pushes', 'pushing' and 'pushed').

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 April; accepted 17 August 2007.

1. Kruskal, J. B., Dyen, I. & Black, P. D. in *Mathematics in the Archaeological and Historical Sciences* (eds Hodson, F. R., Kendall, D. G. & Tautu, P.) 361–380 (Edinburgh Univ. Press, Edinburgh, UK, 1971).
2. Leech, G., Rayson, P. & Wilson, A. *Word Frequencies in Written and Spoken English: based on the British National Corpus* (Longman, London, 2001).
3. Davies, M. Corpus del Español. ⟨http://www.corpusdelespanol.org⟩ (2001–02).
4. Sharoff, S. in *Corpus Linguistics Around the World* (eds Archer, D., Wilson, A. & Rayson, P.) 167–180 (Rodopi, Amsterdam, 2005).
5. Institute for Language and Speech Processing (ILSP) Corpus. Hellenic National Corpus (HNC) Web Version 3.0 [in Greek]. ⟨http://hnc.ilsp.gr/en/⟩ (1999–2006).
6. Gray, R. D. & Atkinson, Q. D. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426,** 435–439 (2003).
7. Gordon, R. G. *Ethnologue: Languages of the World* 15th edn (SIL International, Dallas, 2005).
8. Swadesh, M. Lexico-statistic dating of prehistoric ethnic contacts. *Proc. Am. Phil. Soc.* **96,** 453–463 (1952).
9. Dyen, I., James, A. T. & Cole, J. W. L. Language divergence and estimated word retention rate. *Language* **43,** 150–171 (1967).
10. Pagel, M. & Meade, A. in *Phylogenetic Methods and the Prehistory of Languages* (eds Clackson, J., Forster, P. & Renfrew, C.) 173–182 (MacDonald Institute for Archaeological Research, Cambridge, UK, 2006).

11. Labov, W. *Principles of Linguistic Change: Social Factors* (Blackwell, Oxford, UK, 2001).
12. Milroy, J. & Milroy, L. Linguistic change, social network and speaker innovation. *J. Linguist.* **21,** 229–284 (1985).
13. Nettle, D. Is the rate of linguistic change constant? *Lingua* **108,** 119–136 (1999).
14. Thomason, S. G. & Kaufman, T. *Language Contact, Creolization, and Genetic Linguistics* (Univ. California Press, Berkeley, 1988).
15. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, UK, 1983).
16. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (Univ. Chicago Press, Chicago, 1985).
17. Kirby, S. *Function, Selection, and Innateness: the Emergence of Language Universals* (Oxford Univ. Press, Oxford, UK, 1999).
18. Croft, W. *Explaining Language Change: an Evolutionary Approach* (Longman, Harlow, UK, 2000).
19. Gimbutas, M. The beginning of the Bronze Age in Europe and the Indo-Europeans 3500–2500 B.C. *J. Indo-Eur. Stud.* **1,** 163–214 (1973).
20. Renfrew, C. *Archaeology and Language: the Puzzle of Indo-European Origins* (Cape, London, 1987).
21. Anderson, J. R. Acquisition of cognitive skill. *Psychol. Rev.* **89,** 369–406 (1982).
22. Ellis, N. C. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Stud. Second Lang. Acquisit.* **24,** 143–188 (2002).
23. Fontanari, J. F. & Perlovsky, L. I. Solvable null model for the distribution of word frequencies. *Phys. Rev. E* **70,** 042901 (2004).
24. Zipf, G. K. Prehistoric 'cultural strata' in the evolution of Germanic: The case of Gothic. *Mod. Lang. Notes* **62,** 522–530 (1947).
25. Francis, W. N., Kuçera, H. & Mackie, A. W. *Frequency Analysis of English Usage: Lexicon and Grammar* (Houghton Mifflin, Boston, 1982).
26. Burger, J., Kirchner, M., Bramanti, B., Haak, W. & Thomas, M. G. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl Acad. Sci. USA* **104,** 3736–3741 (2007).
27. Mackay, W. & Kondrak, G. in *Proceedings of the 9th Conf. on Computational Natural Language Learning (CoNLL)* 40–47 (ACL, Schroudsburg, PA, 2005).
28. Greenberg, J. H. *Indo-European and its Closest Relatives: The Eurasiatic Language Family* Vol. 1, *Grammar* (Stanford Univ. Press, Stanford, CA, 2000).
29. Kaiser, M. & Shevoroshkin, V. Nostratic. *Annu. Rev. Anthropol.* **17,** 309–329 (1988).
30. Pagel, M. & Meade, A. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* **53,** 571–581 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Gray and S. Greenhill for comments and advice. This research was supported by a grant to M.P. from the Leverhulme Trust.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.P. (m.pagel@reading.ac.uk).

METHODS

Cognate data. We used the comparative Indo-European database[31], which records word forms and cognacy judgements in 95 languages across the 200 terms in the fundamental vocabulary of ref. 8. We excluded 11 of the speech varieties that had not been coded in ref. 31 and were identified by those authors as less reliable, leaving 84 languages. We added cognacy judgements for the same 200 meanings for three extinct Indo-European languages (Hittite, Tocharian A and Tocharian B), based on multiple sources, for a combined sample of 87 languages[6]. Meanings had from between 1 and 46 cognate sets across the 87 languages, for a total of 4,049 cognates, including unique words.

Phylogenetic trees. We inferred the posterior distribution of phylogenetic trees for the 87 languages using the MCMC methods[32] implemented in *BayesPhylogenies*[30,33]. The cognacy data were transformed to a binary matrix, with rows representing the 87 languages and columns identifying the presence ('1') or absence ('0') of each of the 4,049 cognates (including unique 'cognates'). We then characterized the probability of these data on phylogenetic trees, using a two-state (presence/absence) continuous-time Markov transition rate model[10]. The probability of the data D given the model of evolution M and a tree T is written in the usual way as $P(D|M,T) = \Pi_c P(D|\mathbf{Q}_b,T)$, where here the model of evolution for the binary vectors $\mathbf{Q}_b$ is the 2×2 matrix recording the rates of transition between the binary elements corresponding to the gain of the cognate class and the loss of the cognate class, and the product is over the $c = 4{,}049$ binary vectors that identify cognate classes. The elements of $\mathbf{Q}_b$ are given by $q\pi_j$, where j represents the state (presence/absence) to which the cognate is moving and π is the equilibrium frequency of the jth state[10,34]. We used a single rate parameter q, and estimated the equilibrium frequencies of presence and absence from the data as part of the Markov chain. Transition rates were allowed to vary among cognate class vectors (successive vectors of the binary matrix) according to a gamma distribution[35] with four rate categories. The shape parameter of the gamma distribution was estimated from the data.

We derived the posterior sample of trees from a Markov chain allowed to run for 40,000,000 generations. After discarding the first 2,500,000 generations as burn-in, we sampled every 50,000th tree in the chain to ensure that successive trees were statistically independent. This produced a posterior sample of 750 trees. Examination of autocorrelation times of the MCMC plots indicated that runs had converged to the equilibrium distribution and showed very low autocorrelation, yielding an effective sample size of at least 500. The consensus tree of this posterior sample is reported in Supplementary Fig. 1.

Rates of lexical replacement. We categorized the words for each of the 200 meanings into k states representing the k different cognate classes identified for that meaning. In the meaning class 'big', for example, $k = 24$ because this meaning is represented by 24 cognates among the 87 languages. The probability of observing the distribution of the k lexical terms for a meaning m on any given tree can be written as $P(m|\mathbf{Q}_m,T)$ where $\mathbf{Q}_m$ is a $k \times k$ matrix of the transition rates from any cognate class i to any other class j for a particular meaning, and T is the phylogenetic tree[10]. The elements of $\mathbf{Q}_m$ are given by $q\pi_j$, where q is the instantaneous transition rate (as above but now for a particular meaning) and π_j is the equilibrium frequency of state j (ref. 10). We estimate $\mathbf{Q}_m$ using a continuous-time Markov transition rate model[10,30,33,34]. The equilibrium frequencies are not known and therefore must either be estimated from the data or fixed at prior values. For the results reported in Figs 1 and 3, we assumed uniform equilibrium frequencies across cognates. Thus for a meaning represented by $k = 3$ cognates, each π_j is set to 1/3. Because of the large number of cognate classes for some meanings ($k = 46$ for the meaning 'dirty'), it is impractical to estimate π_j from the data. However, using the observed empirical frequencies of the k classes for each meaning across the 87 languages gives the same qualitative results reported in Fig. 3, and none of our conclusions is altered.

The posterior distribution of the rate parameter q is estimated from a Markov chain that simultaneously proposes new values for q and samples new trees from the posterior distribution of T. We used the mean of the posterior distribution of q to estimate the rates of lexical replacement reported in Figs 1 and 3. Our approach accounts for uncertainty in the model of evolution and in the phylogenetic tree, and, unlike earlier lexicostatistical approaches to estimating rates of cognate replacement, does not suffer from information loss owing to the conversion of cognate data to pair-wise similarity scores between languages[36]. Mean and variance for the 200 meaning category rate estimates are provided in Supplementary Table 2. Rate estimates are not expected to be biased across parts of speech by the process of grammaticalization (see Supplementary Information for details).

Half-life estimates were calculated from the mean q values by solving $P = e^{-qt}$ for t, setting $P = 0.5$ as described in ref. 10.

Word frequency data. Word-use frequencies were obtained for English, Spanish, Russian and Greek from the corpora databases described in refs 2–5. Word frequencies were compiled by searching for all forms listed under the canonical form (or lemma) of each meaning. For example, for the verb meaning 'push' in English, we include 'push', 'pushes', 'pushing' and 'pushed'. Word frequency and part of speech data are provided in Supplementary Table 2.

31. Dyen, I., Kruskal, J. B. & Black, P. An Indo-European classification, a lexicostatistical experiment. 1. *Trans. Am. Phil. Soc.* **82,** 1–132 (1992).
32. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21,** 1087–1091 (1953).
33. Pagel, M. & Meade, A. in *Mathematics of Evolution and Phylogeny* (ed. Gascuel, O.) 121–139 (Oxford Univ. Press, New York, 2005).
34. Pagel, M. & Meade, A. in *The Evolution of Cultural Diversity: a Phylogenetic Approach* (eds Mace, R., Holden, C. J. & Shennan, S.) 235–256 (UCL Press, London, 2005).
35. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39,** 306–314 (1994).
36. Steel, M. A., Hendy, M. D. & Penny, D. Loss of information in genetic distances. *Nature* **336,** 118 (1988).

LETTERS

Phagocyte-derived catecholamines enhance acute inflammatory injury

Michael A. Flierl[1], Daniel Rittirsch[1], Brian A. Nadeau[1], Anthony J. Chen[1], J. Vidya Sarma[1], Firas S. Zetoune[1], Stephanie R. McGuire[1], Rachel P. List[1], Danielle E. Day[1], L. Marco Hoesel[1], Hongwei Gao[1], Nico Van Rooijen[3], Markus S. Huber-Lang[4], Richard R. Neubig[2] & Peter A. Ward[1]

It is becoming increasingly clear that the autonomic nervous system and the immune system demonstrate cross-talk during inflammation by means of sympathetic and parasympathetic pathways[1,2]. We investigated whether phagocytes are capable of *de novo* production of catecholamines, suggesting an autocrine/paracrine self-regulatory mechanism by catecholamines during inflammation, as has been described for lymphocytes[3]. Here we show that exposure of phagocytes to lipopolysaccharide led to a release of catecholamines and an induction of catecholamine-generating and degrading enzymes, indicating the presence of the complete intracellular machinery for the generation, release and inactivation of catecholamines. To assess the importance of these findings *in vivo*, we chose two models of acute lung injury. Blockade of α_2-adrenoreceptors or catecholamine-generating enzymes greatly suppressed lung inflammation, whereas the opposite was the case either for an α_2-adrenoreceptor agonist or for inhibition of catecholamine-degrading enzymes. We were able to exclude T cells or sympathetic nerve endings as sources of the injury-modulating catecholamines. Our studies identify phagocytes as a new source of catecholamines, which enhance the inflammatory response.

Although *tumor*, *rubor*, *dolor* and *calor* (the Latin terms for swelling, redness, pain and heat) are classical features of acute inflammation, the proximal trigger of these responses is neuronal in origin. Similarly, the systemic acute-phase response to infection involves neurone-dependent responses (fever and activation of the hormonal stress response), which are regulated mainly by the hypothalamus[2]. Innate immunity and inflammation are the first lines of defence[4], with initial responses involving non-specific cellular and humoral pathways, resulting in mediator release[1]. Immune and pro-inflammatory mediators that are subsequently released rapidly activate neuronal responses that amplify local immune and inflammatory responses designed to contain pathogens and trigger systemic neuroendocrine and regional neural responses that seek to return the system to a homeostatic state[2]. One of the key pathways involved in the neuroendocrine–immune modulating network is the autonomic nervous system. Recently, vagal parasympathetic signalling has been shown to have an important regulatory function in inflammation through cholinergic receptors on phagocytic cells[1,5,6]. Functional interplay of the adrenergic nervous system with the immune/inflammatory system may counterbalance effects of the parasympathetic nervous system. The presence and synthesis of catecholamines in lymphocytes were first described more than a decade ago[3,7]. Catecholamines can thereby modulate the proliferation, differentiation and apoptosis of lymphocytes and the production of cytokines through adrenoreceptors expressed on T and B cells[8–12]. To investigate whether macrophages and neutrophils (polymorphonuclear cells; PMNs) might be able to generate and release catecholamines, as has been suggested[13,14], cells were incubated with bacterial lipopolysaccharide (LPS) and levels of noradrenaline and adrenaline in cell supernatants were determined. As shown in Fig. 1a–d, noradrenaline and adrenaline levels increased significantly in cell supernatants 15 min after exposure of macrophages and PMNs to LPS, decreasing at 1 h and 2 h, and then increasing by 4 h. Experiments were repeated with human PMNs exposed to the potent complement anaphylatoxin C5a (10 nM), with virtually identical results (data not shown). On the basis of these findings, we conducted experiments to evaluate the presence of two key enzymes involved in catecholamine synthesis in macrophages and PMNs after stimulation with LPS. Tyrosine hydroxylase (TH) is known to be the rate-limiting step in catecholamine synthesis, whereas dopamine β-hydroxylase (DBH) accounts for the final step of noradrenaline synthesis, converting dopamine to noradrenaline[15]. Macrophages or blood PMNs were incubated for 15 min and 4 h with 30 ng ml^{-1} LPS. Thereafter, messenger RNA expression for catecholamine-generating enzymes was analysed by real-time polymerase chain reaction (PCR). Low baseline mRNA levels for both key enzymes (TH and DBH) were detected in unstimulated macrophages. At 4 h after cell exposure to LPS, mRNAs for both enzymes were significantly upregulated (Fig. 1e, f). PMNs expressed low constitutive levels of mRNA for TH and DBH, but both were clearly upregulated 4 h after the addition of LPS (Fig. 1g, h). The actions of catecholamines can be terminated in three ways: by reuptake into nerve terminals, by diffusion into extracellular fluids or by metabolic transformation. Two enzymes are essential in the metabolic inactivation of catecholamines: monoamine oxidase (MAO) and catechol *O*-methyltransferase (COMT). Unstimulated macrophages contained mRNA for both COMT and MAO-A (Fig. 2a, b). After exposure of cells to LPS, the quantity of mRNA for COMT decreased whereas that for MAO-A increased. In accordance with this, when proteins were analysed by western blotting, exposure to LPS caused a decrease in macrophage COMT (Fig. 2c), whereas the same intervention brought about higher levels of MAO-A protein (Fig. 2d). Similar patterns were found in PMNs. Incubation with LPS decreased the level of mRNA for COMT (Fig. 2e) and increased that of mRNA for MAO-A (Fig. 2f). COMT protein amounts decreased when PMNs were exposed to LPS (Fig. 2g); the opposite was true for MAO-A (Fig. 2h). During COMT-mediated *O*-methylation of catecholamines, *S*-adenosylmethionine (SAM) serves as the methyl donor and is subsequently converted to *S*-adenosyl-L-homocysteine (SAH) after donation of the methyl group to the substrate. It has been known for many years that elevated SAH levels can serve as a strong inhibitor of the COMT-mediated *O*-methylation metabolism of catechols[16,17]. Increased phagocyte

[1]Department of Pathology, [2]Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [3]Department of Cell Biology and Immunology, Vrije Universiteit, 1081BT Amsterdam, The Netherlands. [4]Department of Trauma-, Hand- and Reconstructive Surgery, University of Ulm Medical School, 89075 Ulm, Germany.

production of catecholamines might therefore lead to an increased turnover of COMT, increasing levels of the powerful feedback inhibitor SAH and ultimately causing the downregulation of cellular mRNA and protein levels of COMT (Fig. 2). Collectively, these data indicate that LPS-stimulated macrophages and PMNs express catecholamine-generating and degrading enzymes and release noradrenaline and adrenaline after exposure to LPS *in vitro*. Thus, phagocytes possess the complete cellular machinery to generate catecholamines *de novo* and to release and inactivate them.

To evaluate the relevance of these findings *in vivo*, we used two different models of acute lung injury (induced by LPS and by IgG immune complex) to study the effects of adrenoceptor blockade in the setting of a single-organ injury. Adrenoceptors were blocked *in vivo* by using the appropriate antagonists in the two different models of lung inflammatory injury (Fig. 3a, c). The antagonists were delivered intraperitoneally. Vascular leakage of albumin into lungs, which correlates with lung injury, was measured after the deposition of IgG immune complexes (IC) or LPS in the lung. α_1-Adrenoceptors were blocked by prazosin, whereas α_2-adrenoceptors were blocked with RX 821002. β_1-Adrenoceptors were blocked by metoprolol, and β_2-adrenoceptor blockade was induced by ICI 118551. Only blockade of the α_2-adrenoceptors significantly reduced the intensity of the albumin leak into lung in the two models of injury, whereas blockade of other receptors neither intensified nor reduced the level of lung injury (Fig. 3a, c). Blockade of the β_2-adrenoceptors with different doses of ICI 118551 (10–20 mg kg^{-1}) failed to alter injury intensity significantly (data not shown). To clarify the significance of the α_2-adrenoceptors in our models of lung injury, an α_2-adrenergic agonist (UK 14304) was used to assess effects on lung injury. As shown in Fig. 3b, d, the severity of lung damage increased twofold and fourfold, respectively, when induction of IgG or LPS lung injury was followed immediately by intraperitoneal administration of UK 14304, whereas the α_2-adrenoceptor inhibitor RX 821002 substantially suppressed lung injury in each of the two models. The effects of these interventions were consistent with the morphological evaluation of lungs (Fig. 3e–k). Although the levels of lung injury with LPS and IC (positive controls) were comparable (Fig. 3f, i, respectively), animals treated with UK 14304 showed clearly increased evidence of injury as reflected in increased PMN numbers, fibrin and haemorrhage in the interstitial and alveolar compartments (Fig. 3g, j). In sharp contrast, blockade of the α_2-adrenoceptor by RX 821002 reduced the intensity of inflammation and injury (Fig. 4h, k) in comparison with positive controls, indicating a key function for the

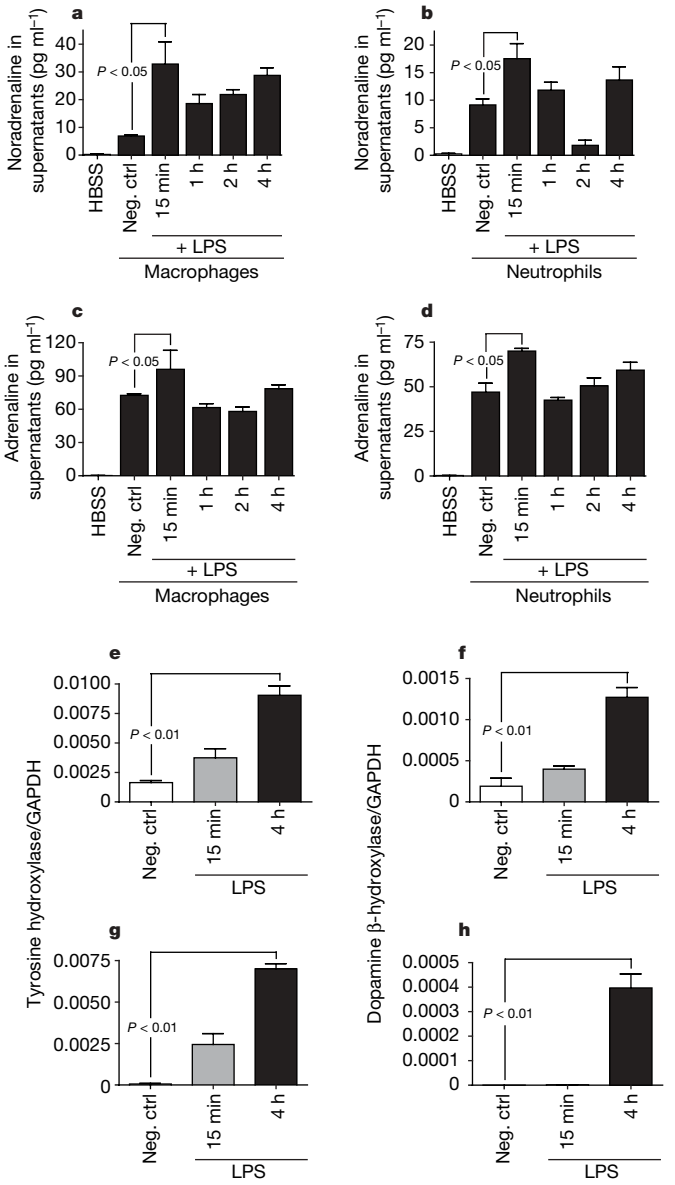


Figure 1 | Release of catecholamines from phagocytes and the presence of catecholamine-producing enzymes in phagocytes. **a**–**d**, After isolation, alveolar macrophages and blood PMNs were stimulated with 30 ng ml^{-1} of LPS *in vitro*. Cellular supernatant fluids were collected as a function of time thereafter and then analysed by enzyme-linked immunosorbent assay for noradrenaline (**a**, **b**) and adrenaline (**c**, **d**). HBSS served as negative control (neg. ctrl). **e**–**h**, After stimulation with 30 ng ml^{-1} of LPS, mRNA from alveolar macrophages (**e**, **f**) and neutrophils (**g**, **h**) was isolated and subjected to real-time PCR analysis for tyrosine hydroxylase and dopamine β-hydroxylase. Each bar represents $n \geq 6$ samples. All data are presented as means ± s.e.m.

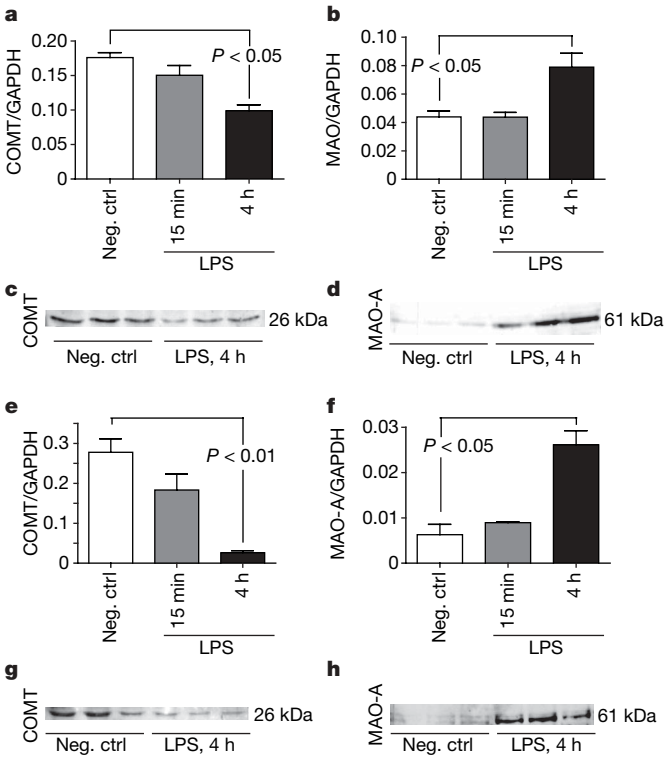


Figure 2 | Expression of catecholamine-inactivating enzymes in phagocytes. After isolation of macrophages and PMNs and stimulation *in vitro* with 30 ng of LPS, mRNA and protein from alveolar macrophages (**a**–**d**) and blood PMNs (**e**–**h**) were subjected to analysis by real-time PCR (**a**, **b**, **e**, **f**) and western blotting (**c**, **d**, **g**, **h**) for the catecholamine-inactivating enzymes COMT (**a**, **c**, **e**, **g**) or monoamine oxidase A (MAO-A) (**b**, **d**, **f**, **h**). Western blots were repeated at least three separate times. Representative blots are shown. Real-time PCR data are presented as means ± s.e.m.; $n \geq 5$ per bar. Neg. ctrl, negative control.

α_2-adrenoceptor in these models of lung injury. α_2-Adrenoceptors belong to the G-protein-coupled, seven-transmembrane receptor family[18] and are currently divided into three subtypes: α_{2A}, α_{2B} and α_{2C} (ref. 19). In the present study we found that both alveolar macrophages and blood PMNs express all three subtypes of the α_2-adrenoceptors (Supplementary Fig. 1).

To exclude T cells or the sympathetic nerve endings as sources of the injury-modulating catecholamines, we achieved T-cell depletion with an antibody and sympathetic denervation by administering reserpine. Recovery of sympathetic function after administration of reserpine takes up to several weeks[20]. In the present study we harvested and exposed PMNs and alveolar macrophages to LPS 4 days after administration of reserpine *in vivo* to determine whether the production of catecholamines by phagocytes is impaired by reserpine. There was no difference in noradrenaline production between phagocytes from control and reserpine-treated animals (Fig. 4a). Accordingly, we induced lung injury in the presence or absence of α_2-adrenoceptor blockade in control animals and animals treated 4 days earlier with reserpine. Injury levels were significantly increased in reserpine-treated animals, whereas α_2-adrenoceptor blockade completely abolished vascular leakage (Fig. 4c). We repeated these experiments *in vitro* and *in vivo* 2 h after reserpine administration, with identical results (data not shown). As expected, adrenal gland homogenates from reserpine-treated animals had almost no detectable noradrenaline (Fig. 4b). In parallel experiments, mice were depleted of CD4 T cells and the intensity of albumin leakage was assessed. There was no change in the level of injury between T-cell-depleted and isotype-control-treated littermates. In the same animals, α_2-adrenoceptor blockade was protective (Fig. 4d). Bronchoalveolar lavage (BAL) fluid analysis was performed to determine whether the sympathetic nervous system and T cells were key sources of catecholamines in current models of lung injury. Peak levels of adrenaline occurred in BAL fluids 2 h after the initiation of immune-complex-mediated injury (Fig. 4e). In parallel, an adrenaline peak was found 3 h after the initiation of LPS-induced lung injury (data not shown). Other mice were depleted of PMNs, alveolar macrophages or T cells or were pharmacologically sympathectomized and subjected to lung injury. BAL fluids were obtained and measured 2 h after the initiation of injury. Whereas depletion of PMNs and alveolar macrophages significantly decreased the presence

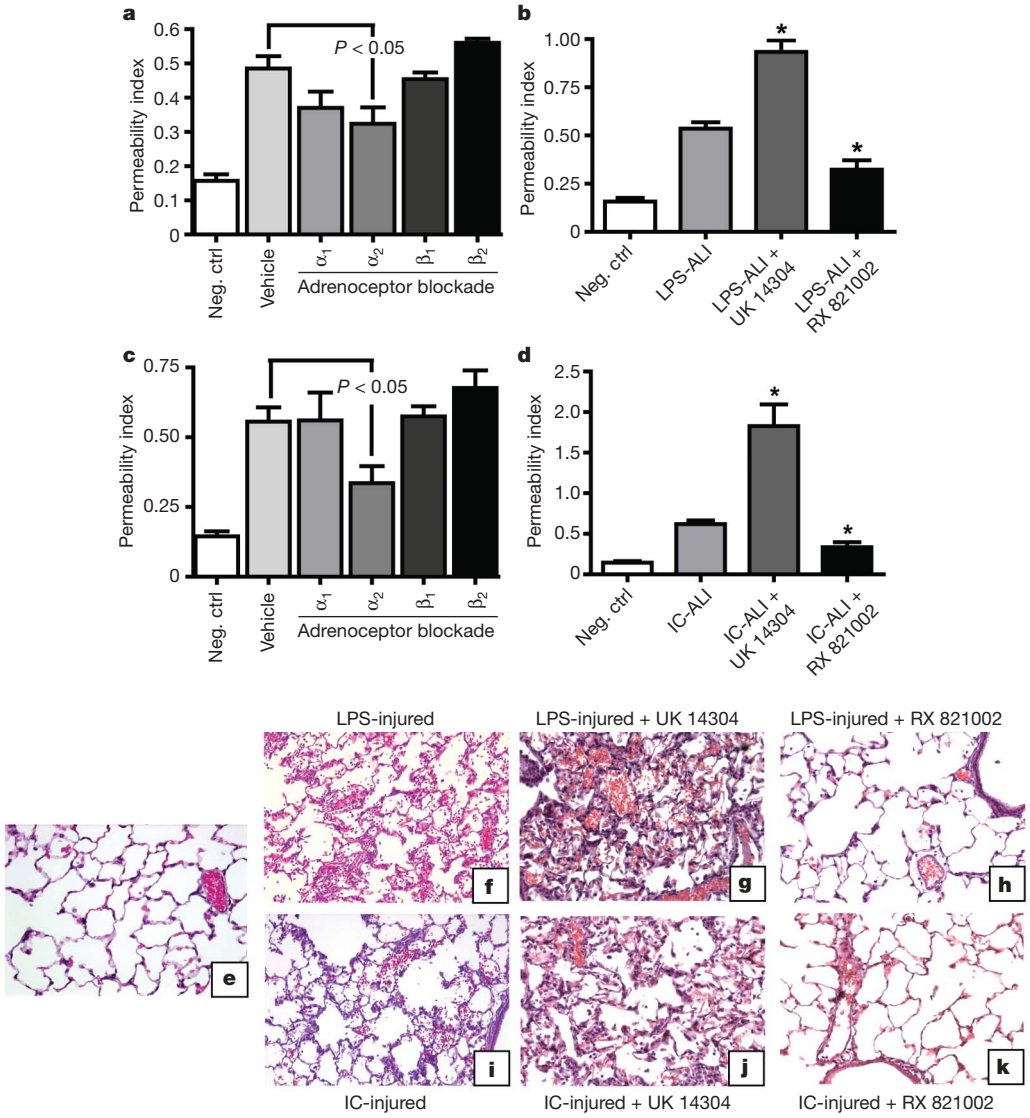


Figure 3 | The α_2-adrenoceptor mediates the severity of experimental acute lung injury. **a**–**d**, Acute lung injury (ALI) was induced by LPS (**a**, **b**) or IgG IC (**c**, **d**) in rats and injury severity was assessed by vascular leak of vascular ^{125}I-albumin. Adrenoceptors were blocked pharmacologically as described in the text. All data are presented as means ± s.e.m.; $n \geq 5$ per bar. Asterisks in **b** and **d** indicate $P < 0.01$ versus LPS-induced and IC-induced lung injury, respectively. **e**–**k**, Hematoxylin and eosin-stained lung sections. **e**, Uninjured lung. **f**–**h**, LPS-injured animals. **i**–**k**, IC-injured littermates. Sections are representative for $n \geq 3$ per group. Neg. ctrl, negative control.

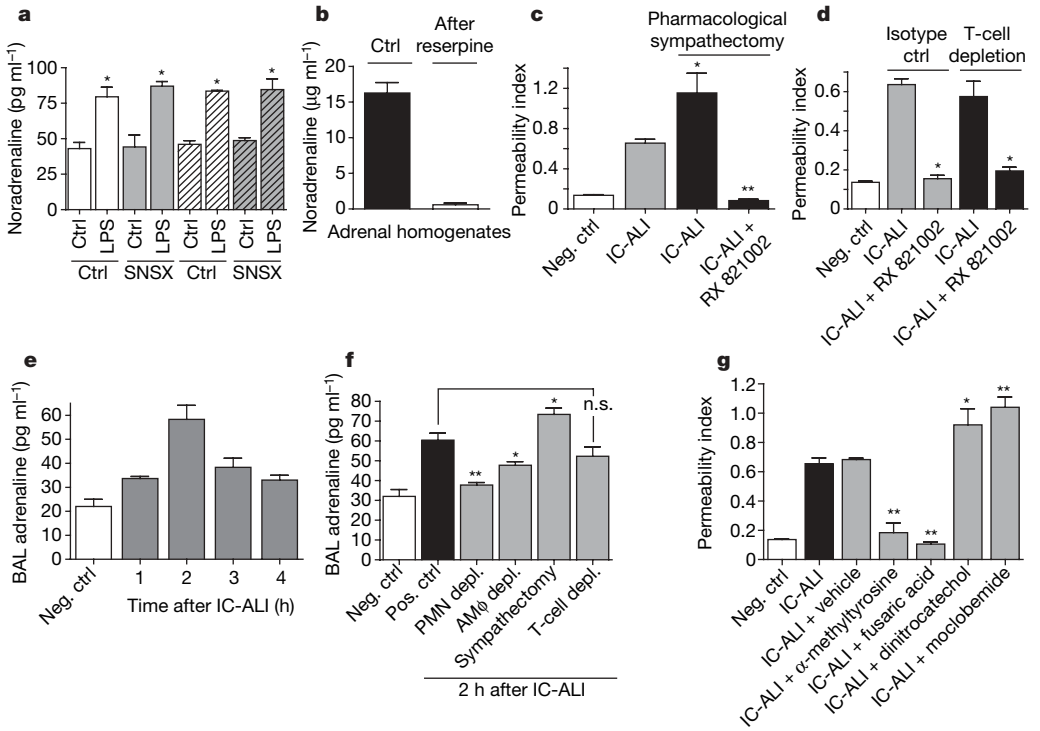


Figure 4 | Phagocyte-derived catecholamines mediate lung injury. a, After sympathectomy (SNSX), noradrenaline production in supernatants of LPS-stimulated phagocytes was assessed. Open bars, PMNs; hatched bars, alveolar macrophages. Asterisk, $P < 0.01$ versus respective control (ctrl). **b**, Sympathectomy was confirmed by adrenal gland homogenates. **c**, Severity of IC-induced acute lung injury (IC-ALI) in sympathectomized animals. Asterisk, $P < 0.05$ versus IC-ALI; two asterisks, $P < 0.01$ versus IC-ALI. **d**, Mice were T-cell depleted, and lung injury was compared with that in isotype-antibody-treated littermates. Asterisk, $P < 0.01$ versus IC-ALI. **e**, Adrenaline peak was determined in BAL fluids. **f**, Mice were depleted (depl.) of PMN, alveolar macrophages (AMΦ) or T cells or were sympathectomized, and adrenaline was measured 2 h after IC-ALI. Asterisk, $P < 0.05$ versus positive control (pos. ctrl); two asterisks, $P < 0.01$ versus positive control. Neg. ctrl, negative control. **g**, Severity of lung injury after inhibition of TH, DBH, COMT or MAO. Asterisk, $P < 0.05$ versus IC-ALI plus vehicle; two asterisks, $P < 0.01$ versus IC-ALI plus vehicle. Data are presented as means ± s.e.m.; $n \geq 5$ per bar.

of adrenaline in BAL fluids, reserpine-induced sympathectomy did not decrease adrenaline levels (Fig. 4f). These data indicate that the presence of adrenaline in BAL fluids requires PMNs and macrophages but not T cells. Phagocytes of adrenalectomized animals have higher baseline levels of TH, DBH and catecholamine secretion (M.A.F., D.R., R.P.L., A.J.C., D.E.D., B.A.N., F.S.Z., J.V.S., A. B. Lentsch, M.S.H.-L. and P.A.W., unpublished observations). It might be possible that the baseline production of catecholamines in phagocytes is increased as a compensatory endogenous mechanism to pharmacological sympathectomy. Last, we showed that pharmacological blockade of the catecholamine-generating TH (by using α-methyltyrosine) and DBH (by using fusaric acid) greatly diminished the inflammatory response, whereas inhibition of the catecholamine-degrading enzymes COMT (with dinitrocatechol) and MAO (with moclobemide) significantly increased lung injury as defined by the permeability index (Fig. 4g).

Alveolar macrophages and PMNs have a central function in the pathophysiology of IgG IC-induced lung injury[21]. The α_2-adrenoreceptor is expressed on alveolar macrophages[11], and modulation of the α_2-adrenoceptor regulates LPS-induced production of tumour necrosis factor (TNF)-α *in vitro* from alveolar macrophages[22]. We sought to determine whether *in vitro* blockade of the adrenoceptors, especially the α_2-adrenoreceptor, might be linked to the present findings *in vivo*. There was a variable inhibition *in vitro* of various macrophage-derived cytokines (TNF-α, interleukin-6 and interleukin-1β) and chemokines (cytokine-induced neutrophil chemoattractant (CINC)-1) by α_1-adrenoceptor, α_2-adrenoceptor, β_1-adrenoceptor or β_2-adrenoceptor blockade (Supplementary Table 1), with blockade of the α_2-adrenoreceptor achieving the most consistent suppression of mediators. Identical experiments with PMNs revealed similar patterns (data not shown).

Taken together, the results of this study indicate that both alveolar macrophages and blood PMNs are capable of the production of catecholamines *de novo* and of their release, implicating phagocytes as a previously unrecognized source of catecholamines. Enhanced production or decreased generation of catecholamines (with the use of inhibitors of catecholamine-generating or catecholamine-degrading enzymes) increases or decreases the intensity of inflammatory injury, respectively. Blockade of α_2-adrenoceptors *in vivo* greatly decreases the intensity of inflammatory injury. In many respects, the phagocytic system can be looked on as a diffusely expressed adrenergic organ.

METHODS SUMMARY

Animals and anaesthesia. All studies were conducted in accordance with the University of Michigan Committee on Use and Care of Animals. Specific pathogen-free male Long–Evans rats (300–325 g; Taconic) or C57BL/6 mice (Jackson Laboratory) were anaesthetized with isoflurane (Aerrane; Henry Schein).

Isolation of PMNs and alveolar macrophages. Whole blood was drawn into syringes containing anticoagulant citrate dextrose (ACD) (Baxter Health Care), in a volume ratio of 1:10. Neutrophils were isolated by using Ficoll-Paque gradient centrifugation (Pharmacia Biotech AB) followed by a dextran sedimentation step. After hypotonic lysis of residual red blood cells, neutrophils were resuspended in Hanks balanced salt solution (HBSS) containing Ca^{2+} and Mg^{2+}. Rat alveolar macrophages were isolated by repeatedly lavaging the lungs of anaesthetized healthy animals. After centrifugation of BAL fluids, cells were resuspended in HBSS containing Ca^{2+} and Mg^{2+}.

Stimulation of phagocytes by LPS. Macrophages or PMNs (3×10^6 or 5×10^6 in 1.0 ml) were isolated and incubated with 30 or 50 ng of LPS (Escherichia coli 0111:B4; Sigma Aldrich). Supernatants were obtained and frozen at −80 °C.

Enzyme-linked immunosorbent assay/enzyme immunoassay analysis of cell supernatants. Noradrenaline and adrenaline enzyme immunoassays were obtained from Rocky Mountain Diagnostics.

Alveolitis induced by IgG IC and by LPS. IgG IC-induced alveolitis was performed as described previously[23,24]. For LPS-induced alveolitis, rats received 300 μg of LPS (Sigma Aldrich) intratracheally.

Morphological assessment of lung injury. At 4 h after IgG IC deposition or 6 h after LPS deposition, lungs were fixed by intratracheal instillation of 10 ml of PBS-buffered (pH 7.2) formalin (10%). The lungs were further fixed in a 10% buffered formalin solution for histological examination by tissue sectioning and staining with haematoxylin and eosin.

Statistical analysis. All values are expressed as means ± s.e.m. Data were analysed with ANOVA and a Student–Newman–Keuls test. Differences were considered significant at $P \leq 0.05$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 July; accepted 20 August 2007.
Published online 30 September 2007.

1. Tracey, K. J. The inflammatory reflex. *Nature* **420,** 853–859 (2002).
2. Sternberg, E. M. Neural regulation of innate immunity: a coordinated nonspecific host response to pathogens. *Nature Rev. Immunol.* **6,** 318–328 (2006).
3. Bergquist, J., Tarkowski, A., Ekman, R. & Ewing, A. Discovery of endogenous catecholamines in lymphocytes and evidence for catecholamine regulation of lymphocyte function via an autocrine loop. *Proc. Natl Acad. Sci. USA* **91,** 12912–12916 (1994).
4. Cook, D. N., Pisetsky, D. S. & Schwartz, D. A. Toll-like receptors in the pathogenesis of human disease. *Nature Immunol.* **5,** 975–979 (2004).
5. Borovikova, L. V. *et al.* Vagus nerve stimulation attenuates the systemic inflammatory response to endotoxin. *Nature* **405,** 458–462 (2000).
6. Wang, H. *et al.* Nicotinic acetylcholine receptor α7 subunit is an essential regulator of inflammation. *Nature* **421,** 384–388 (2003).
7. Musso, N. R., Brenci, S., Setti, M., Indiveri, F. & Lotti, G. Catecholamine content and *in vitro* catecholamine synthesis in peripheral human lymphocytes. *J. Clin. Endocrinol. Metab.* **81,** 3553–3557 (1996).
8. Swanson, M. A., Lee, W. T. & Sanders, V. M. IFN-γ production by Th1 cells generated from naive $CD4^+$ T cells exposed to norepinephrine. *J. Immunol.* **166,** 232–240 (2001).
9. Sanders, V. M. *et al.* Differential expression of the β_2-adrenergic receptor by Th1 and Th2 clones: implications for cytokine production and B cell help. *J. Immunol.* **158,** 4200–4210 (1997).
10. Kohm, A. P., Mozaffarian, A. & Sanders, V. M. B cell receptor- and β_2-adrenergic receptor-induced regulation of B7-2 (CD86) expression in B cells. *J. Immunol.* **168,** 6314–6322 (2002).
11. Bergquist, J., Ohlsson, B. & Tarkowski, A. Nuclear factor-κB is involved in the catecholaminergic suppression of immunocompetent cells. *Ann. NY Acad. Sci.* **917,** 281–289 (2000).
12. Torres, K. C. *et al.* Norepinephrine, dopamine and dexamethasone modulate discrete leukocyte subpopulations and cytokine profiles from human PBMC. *J. Neuroimmunol.* **166,** 144–157 (2005).
13. Brown, S. W. *et al.* Catecholamines in a macrophage cell line. *J. Neuroimmunol.* **135,** 47–55 (2003).
14. Cosentino, M. *et al.* Endogenous catecholamine synthesis, metabolism, storage and uptake in human neutrophils. *Life Sci.* **64,** 975–981 (1999).
15. Elenkov, I. J., Wilder, R. L., Chrousos, G. P. & Vizi, E. S. The sympathetic nerve–an integrative interface between two supersystems: the brain and the immune system. *Pharmacol. Rev.* **52,** 595–638 (2000).
16. Ueland, P. M. Pharmacological and biochemical aspects of *S*-adenosylhomocysteine and *S*-adenosylhomocysteine hydrolase. *Pharmacol. Rev.* **34,** 223–253 (1982).
17. Zhu, B. T. Catechol-*O*-methyltransferase (COMT)-mediated methylation metabolism of endogenous bioactive catechols and modulation by endobiotics and xenobiotics: importance in pathophysiology and pathogenesis. *Curr. Drug Metab.* **3,** 321–349 (2002).
18. Bone, R. C., Francis, P. B. & Pierce, A. K. Intravascular coagulation associated with the adult respiratory distress syndrome. *Am. J. Med.* **61,** 585–589 (1976).
19. Bylund, D. B. *et al.* International Union of Pharmacology nomenclature of adrenoceptors. *Pharmacol. Rev.* **46,** 121–136 (1994).
20. Hoffman, B. B. in *Goodman & Gilman's The Pharmacological Basics of Therapeutics* (eds Brunton, L. L., Lazo, J. S. & Parker, K. L.) 845–868 (McGraw-Hill, New York, 2006).
21. Ward, P. A. Rous–Whipple Award Lecture. Role of complement in lung inflammatory injury. *Am. J. Pathol.* **149,** 1081–1086 (1996).
22. Spengler, R. N., Allen, R. M., Remick, D. G., Strieter, R. M. & Kunkel, S. L. Stimulation of α-adrenergic receptor augments the production of macrophage-derived tumor necrosis factor. *J. Immunol.* **145,** 1430–1434 (1990).
23. Czermak, B. J. *et al.* Synergistic enhancement of chemokine generation and lung injury by C5a or the membrane attack complex of complement. *Am. J. Pathol.* **154,** 1513–1524 (1999).
24. Huber-Lang, M. *et al.* Generation of C5a in the absence of C3: a new complement activation pathway. *Nature Med.* **12,** 682–687 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Schumann and S. Scott for assistance in the preparation of this manuscript. This study was supported by NIH grants to P.A.W. and R.R.N., and by Deutsche Forschungsgemeinschaft grants to M.S.H.-L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.W. (pward@umich.edu).

METHODS

Western blot analysis of MAO-A, COMT, α_{2A}-adrenoceptors, α_{2B}-adrenoceptors and α_{2C}-adrenoceptors. After isolation and exposure to LPS, phagocytes were lysed with Laemmli buffer containing 5% mercaptoethanol. Samples were then separated in a denaturing polyacrylamide gel and transferred to a nitrocellulose membrane. After blocking (5% Tris-buffered saline pH 7.0 containing 0.1% Tween 20) and washing, membranes were then incubated overnight with primary antibodies against adrenoreceptors, MAO-A and COMT (Santa Cruz Biotechnology) and antibodies against GAPDH (Abcam) at 4 °C. Membranes were analysed by enhanced chemiluminescence.

Isolation of mRNA and detection of rat α_{2A}-adrenoreceptors, α_{2B}-adrenoreceptors, α_{2C}-adrenoreceptors, TH, DBH, MAO-A and COMT by quantitative real-time PCR analysis. Total RNA was extracted from phagocytes with Trizol reagent (Life Technologies) in accordance with the manufacturer's instructions. Reverse transcription was performed with 1 μg of RNA by using AMV Reverse Transcriptase (Roche). Real-time quantitative PCR was then performed with SYBR Green PCR Master Mix (Applied Biosystems). The following primer pairs were used: rat α_{2A} adrenoreceptor 5′-CTT GGC CCT CGA CGT GCT CTT TTG-3′ (forward) and 5′-CGG CGG CGC GGA ACA GG-3′ (reverse); rat α_{2B} adrenoreceptor, 5′-TCG GCC ATC ACC TTT CTC ATC CTT-3′ (forward) and 5′-GGG CGT CGG GGC GTT GGT C-3′ (reverse); rat α_{2C} adrenoreceptor, 5′-TGG CTC ATC TCG GCT GTC ATC TCC-3′ (forward) and 5′-TGT CCC CCT CGG CAC CCT CTC-3′ (reverse); rat TH, 5′-AGT CTG GCC TTC CGC GTG TTT CAA-3′ (forward) and 5′-GGG CGC TGG ATA CGA GAG GCA TAG-3′ (reverse); rat DBH, 5′-CTG CGA CCC CAA GGA TTA TG-3′ (forward) and 5′-CAG CAC GTG GCG ACA GTA GTT-3′ (reverse); rat MAO-A, 5′-GTT GAG CGG CTG ATC CAC TTT GT-3′ (forward) and 5′-CCG GCC TCT CCA GCT TCA CTC T-3′ (reverse); rat COMT, 5′-GAC GCG AAA GGC CAA ATC AT-3′ (forward) and 5′-GCT GCT GCT CCC TCT CAC ATA-3′ (reverse); rat GAPDH, 5′-CGG CAA GTT CAA CGG CAC AGT CA-3′ (forward) and 5′-CTT TCC AGA GGG GCC ATC CAC AG-3′ (reverse). Product sizes were 403 base pairs (bp) (α_{2A}-adrenoreceptor), 427 bp (α_{2B}-adrenoreceptor), 409 bp (α_{2C}-adrenoreceptor), 458 bp (TH), 439 bp (DBH), 451 bp (MAO-A), 435 bp (COMT) and 424 bp (GAPDH).

Drugs. All drugs (from Sigma Aldrich unless stated otherwise) were administered intraperitoneally: propane-1,2-diol (vehicle; 1 ml kg^{-1}), Prazosin (2 mg kg^{-1}), RX 821002 (10 mg kg^{-1}), metoprolol (15 mg kg^{-1}), ICI 118551 (10 mg kg^{-1}) and UK 14304 (2 mg kg^{-1}), α-methyltyrosine (200 mg kg^{-1}), fusaric acid (75 mg kg^{-1}), 3,5-dinitrocatechol (50 mg kg^{-1}) and moclobemide (20 mg kg^{-1}; Biotrend).

Depletion of CD4 T cells, neutrophils and alveolar macrophages. For depletion of CD4 T cells, C57BL/6 mice were treated with 0.2 mg of CD4-specific monoclonal antibody (GK1.5; BD Pharmingen) intraperitoneally 48 h before induction of lung injury. Alveolar macrophages and neutrophils were depleted as described previously[25,26].

Pharmacological sympathectomy. To achieve pharmacological sympathectomy, mice were injected intraperitoneally with reserpine (5 mg kg^{-1}). Four days after reserpine injection, acute lung injury was initiated as described above.

25. Gao, H. *et al.* Stat3 activation in acute lung injury. *J. Immunol.* **172**, 7703–7712 (2004).
26. Hoesel, L. M. *et al.* Harmful and protective roles of neutrophils in sepsis. *Shock* **24**, 40–47 (2005).

LETTERS

Mechanism of chloride interaction with neurotransmitter:sodium symporters

Elia Zomot[1], Annie Bendahan[1], Matthias Quick[2,4], Yongfang Zhao[2], Jonathan A. Javitch[2,3,4] & Baruch I. Kanner[1]

Neurotransmitter:sodium symporters (NSS) have a critical role in regulating neurotransmission and are targets for psychostimulants, anti-depressants and other drugs[1,2]. Whereas the non-homologous glutamate transporters mediate chloride conductance[3], in the eukaryotic NSS chloride is transported together with the neurotransmitter[4–7]. In contrast, transport by the bacterial NSS family members LeuT, Tyt1 and TnaT is chloride independent[8–10]. The crystal structure of LeuT reveals an occluded binding pocket containing leucine and two sodium ions[9], and is highly relevant for the neurotransmitter transporters[11–13]. However, the precise role of chloride in neurotransmitter transport and the location of its binding site remain elusive. Here we show that introduction of a negatively charged amino acid at or near one of the two putative sodium-binding sites of the GABA (γ-aminobutyric acid) transporter GAT-1 from rat brain (also called SLC6A1)[14,15] renders both net flux and exchange of GABA largely chloride independent. In contrast to wild-type GAT-1, a marked stimulation of the rate of net flux, but not of exchange, was observed when the internal pH was lowered. Equivalent mutations introduced in the mouse GABA transporter GAT4 (SLC6A11) and the human dopamine transporter DAT (SLC6A3) also result in chloride-independent transport, whereas the reciprocal mutations in LeuT and Tyt1 render substrate binding and/or uptake by these bacterial NSS chloride dependent. Our data indicate that the negative charge, provided either by chloride or by the transporter itself, is required during binding and translocation of the neurotransmitter, probably to counterbalance the charge of the co-transported sodium ions.

In the crystal structure of a chloride/proton antiporter[16] of the ClC family, the chloride ions are coordinated by main-chain NH groups and by side-chain hydroxyls from serine and tyrosine residues[17,18]. To identify potential candidate amino acids contributing to the interaction of the neurotransmitter transporters of the NSS family with chloride, we looked for serine, threonine and tyrosine residues located in the transmembrane domains (TM) that were conserved in these transporters but not in their bacterial counterparts, in which transport is independent of chloride[8–10]. This criterion was met by a threonine–serine pair located in TM 7, just one helical turn 'above' a conserved asparagine (Fig. 1a). In the structure of LeuT from *Aquifex aeolicus* this asparagine coordinates one of the two bound sodium ions (Na1)[9] (Fig. 1b). The positions of the conserved threonine–serine pair in the sodium- and chloride-coupled GABA transporter GAT-1 are 330 and 331, respectively, and the asparagine is at position 327 (Fig. 1a). The threonine–serine pair is not conserved in the bacterial family members LeuT, TnaT (from *Symbiobacterium thermophilum*) and Tyt1 (from *Fusobacterium nucleatum*), but one or more acidic residues are located in the short stretch corresponding to GAT-1 positions 327–331 (Fig. 1a). Mutants of GAT-1 with an acidic amino acid introduced at position 330 were devoid of detectable [^{3}H]GABA transport activity when expressed in HeLa cells[19], but the corresponding mutants at position 331 (S331E and S331D) had significant, albeit reduced, activity (Fig. 1c). In marked contrast to wild-type GAT-1, where transport activity in the absence of chloride (gluconate substitution) was around 1–2% of that in its presence, transport by S331E and S331D was almost completely chloride independent (Fig. 1d). This phenotype was dependent on the introduction of an acidic amino acid at this position, and was not seen with the S331A or the S331G mutants (Fig. 1d), which had similar transport activity compared to S331D (Fig. 1c). Similarly, mutation of Asn 327 to an acidic residue, but not to cysteine, resulted in chloride-independent transport, and the same was true for position 328, but not for position 329 (Fig. 1e, f). The low activity of many of the mutants is probably due to a direct or indirect perturbation of the Na1 site (Supplementary Discussion), which presumably also underlies the altered sodium concentration dependence of transport by S331E (Supplementary Fig. 1). Moreover, the Michaelis constant (K_m) for GABA of the S331E mutant (0.88 ± 0.16 μM and 0.78 ± 0.20 μM, in the presence and absence of chloride, respectively; $n = 4$) was even lower than the corresponding wild-type value (1.35 ± 0.18 μM, could only be measured in the presence of chloride; $n = 5$). However, the maximal initial velocity (V_{max}) of S331E was only 5.8 ± 1.7% of that of wild-type GAT-1.

We proposed that the cause of the reduced V_{max} of the GAT-1 S331E mutant was related to its chloride-independent [^{3}H]GABA transport (Fig. 1d). In the transport cycle of wild-type GAT-1 the coupled translocation of GABA with one chloride and two sodium ions[4–6] is followed by the reorientation of the empty binding sites to enable the next cycle (Fig. 2a). In the S331E mutant, GABA transport is independent of chloride, because apparently the negatively charged side chain of the introduced glutamate fulfils the role of chloride. In contrast to wild-type GAT-1, where the translocated chloride ion is released into the cytoplasm, at physiological pH the engineered glutamate would remain negatively charged in the empty S331E transporter, rendering the reorientation step in the S331E mutant strongly rate-limiting. If this were the case, lowering the intracellular pH would neutralize this negative charge, leading to an increased rate of transport (Fig. 2b). This prediction was tested on reconstitution of wild-type and S331E transporters in liposomes[14,20]. In this system an artificial sodium gradient can be imposed using potassium as internal cation and sodium in the external medium. In the reconstituted system, [^{3}H]GABA transport by wild-type GAT-1 was also almost completely dependent on the presence of chloride in the external medium (Fig. 2c). Addition of chloride only to the internal medium did not support transport (data not shown), and lowering the internal pH from 7.4 to 6.0 did not stimulate transport by wild-type GAT-1 (Fig. 2c). In marked contrast, [^{3}H]GABA transport by

[1]Department of Biochemistry, Hebrew University Hadassah Medical School, POB 12272, Jerusalem 91120, Israel. [2]Center for Molecular Recognition and [3]Departments of Psychiatry and Pharmacology, Columbia University, 630 W. 168th, New York, New York 10032, USA. [4]Division of Molecular Therapeutics, New York State Psychiatric Institute, New York, New York 10032, USA.

liposomes inlaid with the S331E mutant was increased by almost an order of magnitude when the internal pH was lowered to 6.0 (Fig. 2d), and the transport rates were 40–50% of those of wild-type GAT-1. In further contrast to the wild-type protein, very similar transport data were obtained in the absence of chloride (Fig. 2d). This chloride-independent transport by S331E proteoliposomes was still fully dependent on sodium (substitution by Tris) and potently stimulated by the potassium-specific ionophore valinomycin, which can shunt the charge translocated by electrogenic sodium-coupled GABA transport (Supplementary Fig. 2). GABA influx by wild-type GAT-1, but not by the S331E mutant, was potently inhibited by internal chloride. Transport in the presence of 75 mM internal chloride (75 mM KCl replaced 75 mM KMES, conditions of Fig. 2) was 25.5 ± 4.9% of that in its absence for wild-type GAT-1, but 102.8 ± 11.8% for the S331E mutant ($n = 4$, 10 min uptake). The effect of internal chloride on influx by wild-type GAT-1 does not support a proposed mechanism of internal and external chloride exchange during GABA influx[21], but implies chloride and GABA co-transport[4,5,22]. Chloride exchange was proposed because the net inward current during GABA transport seemed to be totally accountable by the influx of sodium ions[21], but recent observations indicate that part of this current is due to a chloride-independent, GABA-gated conduction of sodium ions[6].

Transport in proteoliposomes, where internal potassium is replaced by sodium and unlabelled GABA, reflects exchange of external [^{3}H]GABA with unlabelled internal GABA. This partial

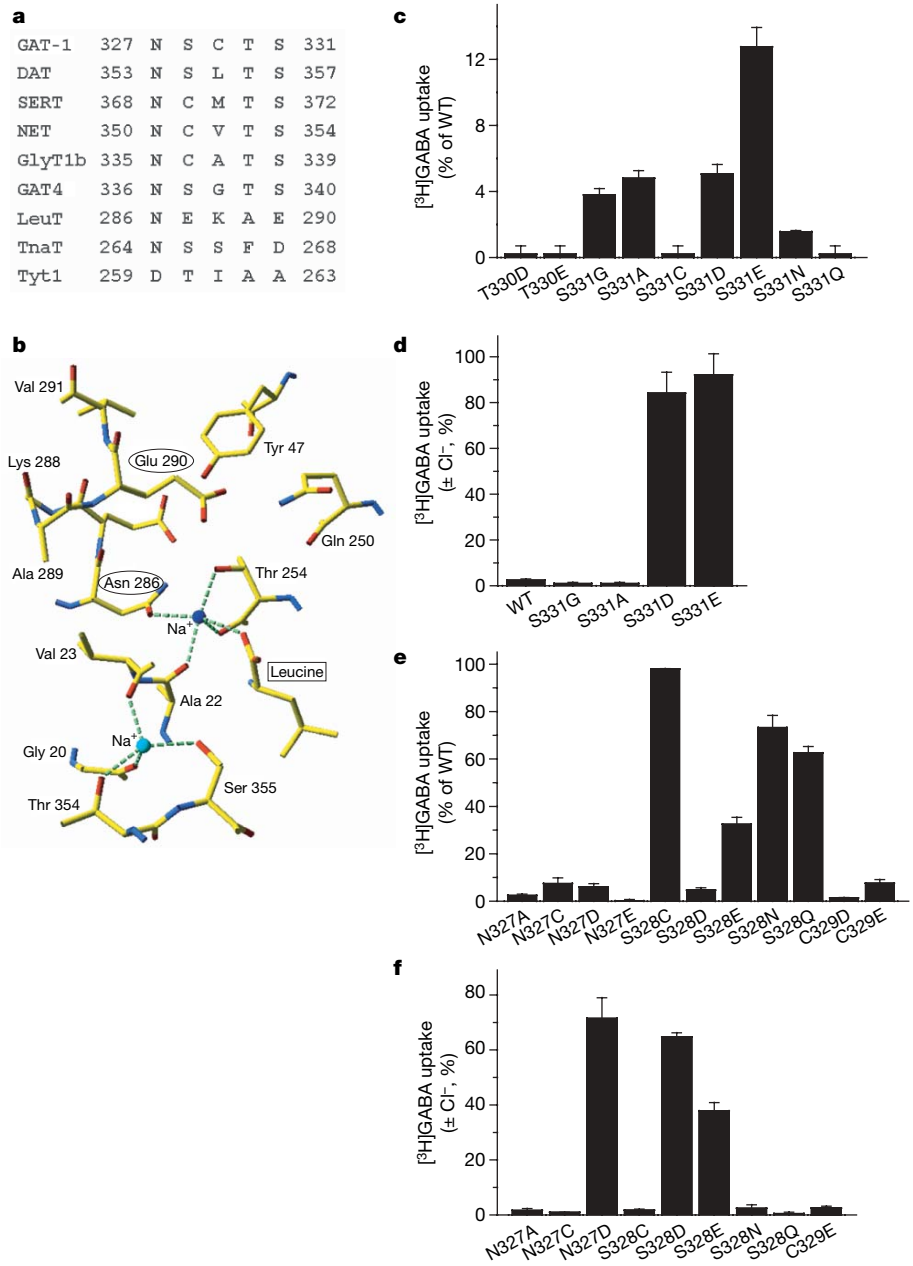


Figure 1 | A putative chloride site in GAT-1. **a**, Amino acid sequence alignment of selected NSS members. The transporters are described in the text, except for the norepinephrine transporter NET and the glycine transporter GlyT1b. **b**, Parts of the LeuT binding pocket (Na1, blue; Na2, cyan; leucine, boxed label) are indicated. Ionic interactions are depicted as dashed green lines. Asn 27 and Ala 351, which interact with Na1 and Na2, respectively, have been omitted for clarity; Asn 286 and Glu 290, the counterparts of GAT-1 Asn 327 and Ser 331, respectively, are highlighted (ellipse). **c**, **e**, [^{3}H]GABA transport by HeLa cells expressing the indicated replacement GAT-1 mutants. **d**, **f**, Chloride dependence (gluconate replacement) of [^{3}H]GABA uptake by wild-type GAT-1 and replacement mutants. $\pm Cl^-$, with or without chloride. The values shown are mean ± standard error of at least three experiments each done in triplicate.

transport reaction does not require the reorientation of the empty transporter (Fig. 2a, b). Exchange by wild-type GAT-1 was still totally dependent on the presence of chloride, regardless of whether the internal pH was 6.0 (Supplementary Fig. 3a) or 7.4 (data not shown). Just like net flux, exchange by the S331E mutant was essentially chloride independent (~90% in the absence of chloride compared with the value in its presence; data not shown). However, in contrast to net flux by the mutant (Fig. 2d), exchange was not stimulated, but rather inhibited, when the internal pH was lowered from 7.4 to 6.0, whether in the absence of chloride (Supplementary Fig. 3b) or in its presence (data not shown). Thus, the potent stimulation of net flux of GABA by the S331E mutant at lowered pH (Fig. 2d) is apparently owing to the acceleration of the return step of the empty transporter. The ratio of the exchange (monitored after 10 min) at internal pH 6.0 compared to that at pH 7.4 was 1.35 ± 0.25 for wild-type GAT-1 ($n = 4$) in the presence of chloride; for the S331E mutant this ratio was 0.85 ± 0.07 and 0.46 ± 0.05 in the presence and absence of chloride, respectively ($n = 4$). The simplest scenario is that during influx, the proton carried by the protonated, but otherwise 'empty', S331E mutant is released to the outside (Fig. 2b); however, the possibility that the proton is returned to the inside can not be excluded.

Our findings, that introduction of a negatively charged amino acid residue near the putative Na1 (ref. 9) site in GAT-1 renders GABA transport independent of chloride, are not limited to this member of the NSS family. Similar results were seen for two other neurotransmitter transporters of this family—the dopamine transporter DAT and the GABA transporter GAT4—when a glutamate residue was introduced at the position corresponding to residue 331 of GAT-1. Transport of [^{3}H]GABA by wild-type GAT4 was strongly dependent on the presence of chloride (gluconate substitution), whereas transport by the S340E GAT4 mutant was largely independent of the presence of this anion in the extracellular medium (Fig. 3a). The same was true for [^{3}H]dopamine transport by wild-type DAT and its S375E mutant (Fig. 3c). Transport by both of these mutants remained sodium dependent, even in the absence of chloride (Fig. 3b, d).

Our work is based on the assumption that the LeuT structure[9] provides a valid structural model for GAT-1 and the other neurotransmitter transporters of the NSS family. We obtained independent support for this idea by making LeuT and Tyt1—two chloride-independent bacterial NSS—chloride dependent (Fig. 3e, f). In the case of LeuT we made the E290S mutant, which is reciprocal to the S331E mutant of GAT-1 (Fig. 1a). Although we were unable to measure L-[^{3}H]alanine transport[23] by the LeuT E290S mutant reconstituted into proteoliposomes, we observed L-[^{3}H]leucine binding to E290S, albeit with reduced apparent affinity for leucine and for sodium (data not shown). Notably, leucine binding to the LeuT E290S mutant in the absence of chloride was only around 12% of that in its presence, in contrast to binding to wild-type LeuT, which was completely chloride independent (Fig. 3e). Because

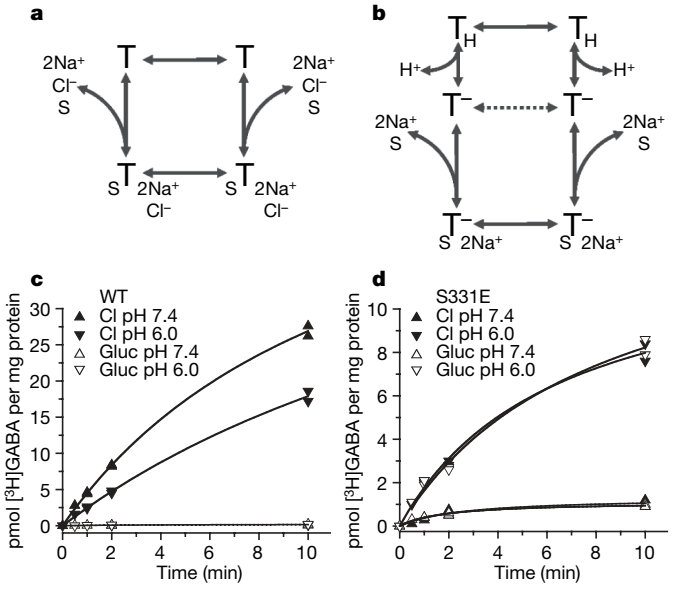


Figure 2 | Net flux by wild-type and chloride-independent mutant transporters: theory and experiment. **a**, Transport cycle of wild-type GAT-1: after binding of two sodium ions, chloride and substrate (S) from the outside of the cell (left), their translocation and release to the inside (right) is followed by the reorientation of the empty transporter (T). **b**, The negative charge of GAT-1 S331E (T^-) enables the translocation of sodium and substrate even in the absence of chloride, but the return step of the empty transporter is slow (dashed, double-headed arrow) unless the negative charge on the transporter is neutralized by protonation (T_H). **c**, **d**, Time course of net [^{3}H]GABA transport by reconstituted wild-type GAT-1 (**c**) or the S331E mutant (**d**). Liposomes were loaded with 0.12 M KPi, pH 7.4 (triangles), or with 0.15 M KMES, pH 6.0 (inverted triangles); the external medium contained 75 mM NaCl plus 60 mM NaPi, pH 7.4 (filled symbols), or 75 mM Na-gluconate plus 60 mM NaPi, pH 7.4 (open symbols). The double symbols represent duplicate measurements.

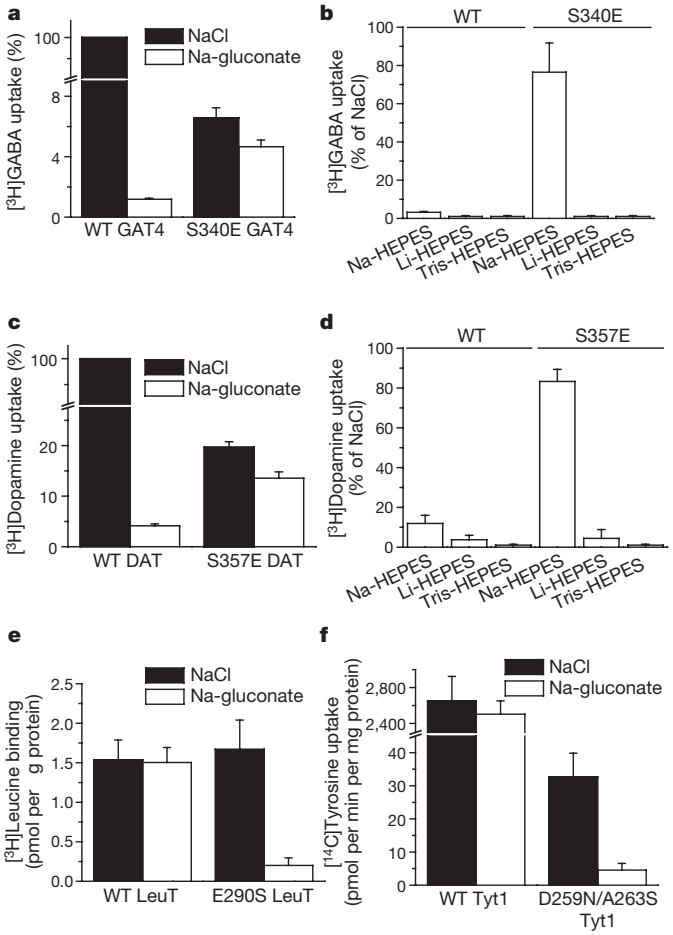
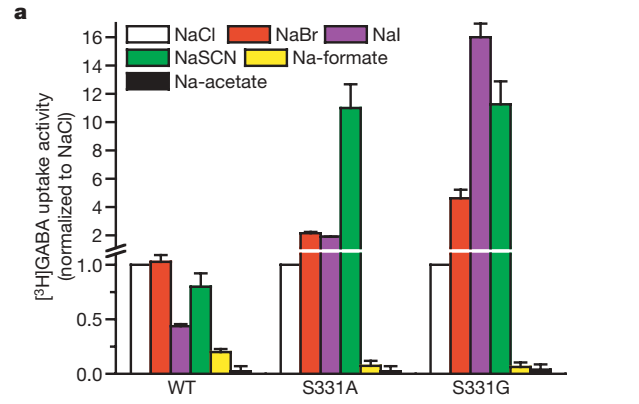
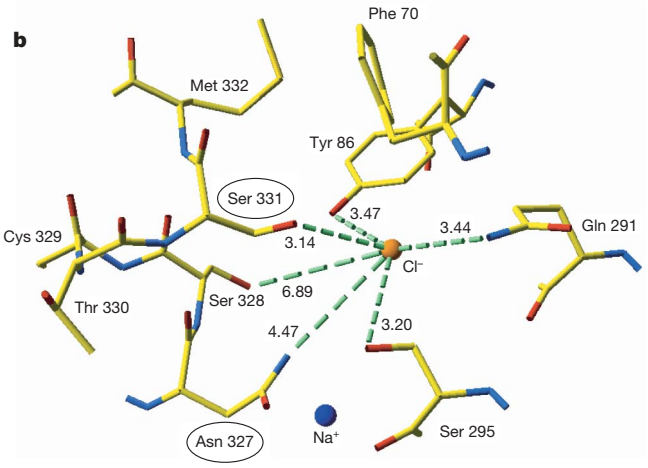
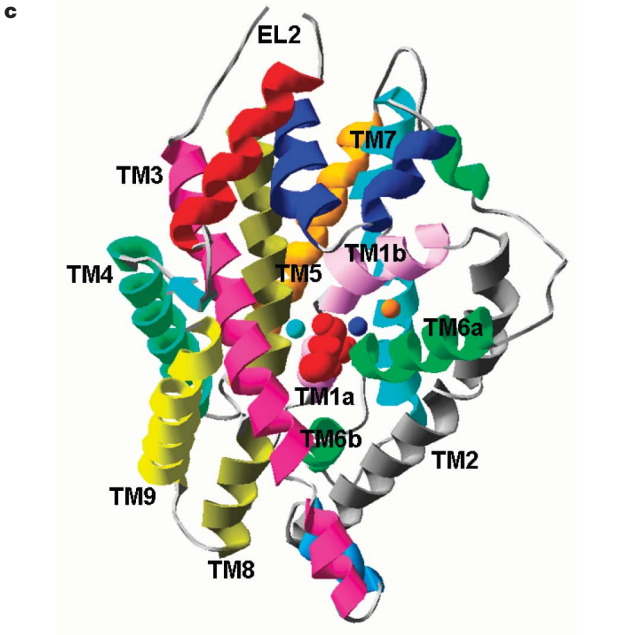


Figure 3 | Transport and binding by other NSS transporters. Measurements were performed in the presence (filled bars) or absence of chloride (gluconate replacement, open bars). The values shown are mean ± s.e. of at least three experiments each done in triplicate. **a**–**d**, Transport of [^{3}H]GABA (**a**, **b**) or [^{3}H]dopamine (**c**, **d**) by HeLa cells, expressing wild-type GAT4 or DAT, or the indicated replacement mutants, was measured as described in Methods. **b**, **d**, Effect of cationic substitution (120 mM final concentration, pH 7.4, adjusted with HEPES) on substrate transport by GAT4 (**b**) or DAT (**d**). **e**, Leucine binding by wild-type LeuT or the E290S mutant was measured by means of scintillation proximity-based assays (see Methods). **f**, Uptake of [^{14}C]tyrosine by intact *E. coli* MQ614 expressing wild-type Tyt1 or the D259N/A263S double mutant in the presence or absence of chloride (see Methods).

transport activity in Tyt1 is much faster than in LeuT, we attempted to create a chloride-dependent Tyt1 mutant. In contrast to LeuT, the corresponding mutation in Tyt1 (A263S) did not alter the chloride-independent phenotype of this transporter (data not shown). Because Tyt1 has Asp 259 at the aligned position of Asn 327 in GAT-1, we also mutated Asp 259 to asparagine or serine. Both of these mutants showed substantially reduced, chloride-independent tyrosine uptake (data not shown). However, uptake by the double mutant Tyt1(D259N/A263S) in the absence of chloride was only around 14% of that in its presence, in contrast to transport by wild-type Tyt1, which was completely chloride independent (Fig. 3f).







If Ser 331 participates directly in the binding of chloride, truncation of its side chain would create more space in the binding site. Indeed, when Ser 331 was mutated to a smaller residue, such as alanine or glycine, in GAT-1, larger anions, such as iodide and thiocyanate, became markedly more effective in their ability to stimulate transport (Fig. 4a). This transport was fully dependent on sodium (Supplementary Fig. 4). To visualize a potential chloride-binding site, we used homology modelling and energy minimization based on the LeuT structure. The result is a cavity where the chloride ion may interact with the hydrogen atoms of the amide group of Gln 291 and of the hydroxyl groups of Ser 331, Ser 295 and Tyr 86 (GAT-1 numbering) (Fig. 4b). In GAT-1, Tyr 86 is on TM 2 and is conserved in almost all members of the NSS family. Replacing Tyr 86 with glutamate or aspartate resulted in transport activity of $50 \pm 4\%$ ($n = 7$) and $16 \pm 4\%$ ($n = 4$) of that of wild type, respectively, and a modest gain in the chloride-independent [^{3}H]GABA transport was observed in both cases (~7% of that in the presence of chloride, data not shown). Introduction of an aspartate or a glutamate at position 295 abolished transport almost completely, apparently because of the role of Ser 295 in the coordination of Na1 (ref. 9). The Q291E mutant is inactive in transport[24], presumably because of the role of Gln 291 in the formation of the extracellular gate[9]. In LeuT, the side chain of the equivalent of Asn 327 also binds Na1[9]. Notably, GAT-1 N327D has chloride-independent transport (Fig. 1f), suggesting proximity of Na1 to the chloride site. Even though in this model Ser 328 appears to be too far away to interact with chloride, the negative charge of the S328E and S328D mutants would be relatively close to the position of the chloride. Indeed, these mutations result in chloride-independent transport (Fig. 1f).

The proximity of the putative chloride and Na1 sites (Fig. 4b, c) and the requirement of the negative charge during the translocation of GABA and the two sodium ions, but not during the return step, suggest that the role of chloride is mainly to compensate the multiple positive charges. This proximity also offers a satisfying explanation for the ability of chloride to increase the apparent affinity of GAT-1 for sodium by chloride[25]. In GAT-1 S331E and in LeuT and Tyt1, the negative charge of the glutamate or aspartate side chain mimics the role of chloride in substrate transport. Indeed, the sodium-concentration dependence of transport by S331E, but not by wild-type GAT-1, was right-shifted when the external pH was lowered from 7.4 to 5.8 (Supplementary Fig. 5). Most other NSS apparently also use multiple sodium ions. The compensation of multiple positive charges by chloride apparently enables the neurotransmitter transporters to accumulate their substrate against huge concentration gradients.

Note added in proof: Following submission of the final version of this manuscript, a paper dealing with the chloride interaction site in the NSS was published[27]. The findings in that paper are in full harmony with our own observations.

Figure 4 | Anion selectivity of GAT-1 S331 mutants and putative chloride-binding site in the NSS. a, [^{3}H]GABA transport by HeLa cells expressing wild-type GAT-1, S331A or S331G. To maintain chloride-free conditions, $CaCl_2$ was replaced by Ca-gluconate and 150 mM of the sodium salts was used as indicated. The relative activity of the two mutants is given in Fig. 1c. The values shown are mean ± standard error of at least three experiments done in triplicate. **b**, **c**, To illustrate schematically the proposed chloride-binding site in eukaryotic NSS (and the bacterial NSS mutants), the chloride ion (orange), Na1 (blue) and Na2 (cyan), and bound leucine (red, space-filling representation) are shown in the context of the overall LeuT structure. **b**, Glu 290 of LeuT was replaced with serine (equivalent to Ser 331 of GAT-1 (Fig. 1a)), a chloride ion (orange sphere) was introduced at the site of the γ-carboxyl group of the original glutamate residue, and residues within 5 Å of the introduced serine were replaced by their GAT-1 counterparts. Energy minimization was applied using Sybyl7.0 software. Numbering refers to GAT-1 and distances are given in ångstroms. TM10–TM12 (**c**) and Asn 66 (roughly positioned 'in front of' Asn 327) (**b**) have been removed for clarity. **b** and **c** were prepared using DeepView Swiss-PDB viewer downloaded from the ExPaSy Proteomics Server.

METHODS SUMMARY

Transporter complementary DNA clones and site-directed mutagenesis. The cDNAs encoding GAT4, DAT and LeuT were gifts from N. Nelson, M. Caron and E. Gouaux, respectively. Site-directed mutagenesis was done using single-stranded uracil containing DNA (mammalian transporters) or by standard polymerase chain reaction (PCR) procedures (bacterial transporters). Mutagenic DNA fragments flanked by unique restriction sites were subcloned back into the wild type using appropriate pairs of restriction enzymes. Modified plasmids were introduced into *Escherichia coli* XL1-Blue and the integrity of the inserted fragments was confirmed by DNA sequencing.

Transporter expression and reconstitution. HeLa cells were infected with recombinant vaccinia virus encoding T7 polymerase, followed by transfection with cationic phospholipids and plasmid DNA containing wild-type or mutant transporters, cloned in the vector pBluescript SK(−) behind the T7 promoter. Transport of [^{3}H]GABA or [^{3}H]dopamine was measured in sodium-containing media. For reconstitution of transporters, HeLa cells expressing wild-type or mutant transporters were treated with the detergent cholate, added to a mixture of phospholipids from soy beans and bovine brain, followed by detergent removal using spin columns. Sodium-dependent transport of [^{3}H]GABA was determined using a filtration assay. Expression of bacterial transporters and uptake in whole cells was performed as described[10]

Determination of transporter levels at the plasma membrane. Transporters in the plasma membrane were labelled with a membrane-impermeant biotinylation reagent, solubilized with detergent and isolated with streptavidin beads. After elution from the beads and SDS–polyacrylamide gel electrophoresis (PAGE), the transporter bands were visualized using an antibody directed against a peptide sequence from the carboxy terminus of GAT-1, a secondary antibody, and ECL.

Substrate binding and transport by bacterial NSS. Substrate binding to wild-type and mutant transporters was monitored using a scintillation proximity assay[26] and transport was assayed as described[10].

Homology modelling. Homology modelling of the putative chloride-binding site was done using the Tripos Sybyl7.0 computational software.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 March; accepted 1 August 2007.
Published online 19 August 2007.

1. Kanner, B. I. Sodium-coupled neurotransmitter transport: structure, function and regulation. *J. Exp. Biol.* **196,** 237–249 (1994).
2. Giros, B., Jaber, M., Jones, S. R., Wightman, R. M. & Caron, M. G. Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. *Nature* **379,** 606–612 (1996).
3. Wadiche, J. I., Amara, S. G. & Kavanaugh, M. P. Ion fluxes associated with excitatory amino acid transport. *Neuron* **15,** 721–728 (1995).
4. Keynan, S. & Kanner, B. I. γ-Aminobutyric acid transport in reconstituted preparations from rat brain: coupled sodium and chloride fluxes. *Biochemistry* **27,** 12–17 (1988).
5. Hilgemann, D. W. & Lu, C. C. GAT1 (GABA:Na^+:Cl^-) cotransport function. Database reconstruction with an alternating access model. *J. Gen. Physiol.* **114,** 459–475 (1999).
6. Krause, S. & Schwarz, W. Identification and selective inhibition of the channel mode of the neuronal GABA transporter 1. *Mol. Pharmacol.* **68,** 1728–1735 (2005).
7. Roux, M. J. & Supplisson, S. Neuronal and glial glycine transporters have different stoichiometries. *Neuron* **25,** 373–383 (2000).
8. Androutsellis-Theotokis, A. *et al.* Characterization of a functional bacterial homologue of sodium-dependent neurotransmitter transporters. *J. Biol. Chem.* **278,** 12703–12709 (2003).
9. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature* **437,** 215–223 (2005).
10. Quick, M. *et al.* State-dependent conformations of the translocation pathway in the tyrosine transporter Tyt1, a novel neurotransmitter:sodium symporter from *Fusobacterium nucleatum*. *J. Biol. Chem.* **281,** 26444–26454 (2006).
11. Zhou, Y., Zomot, E. & Kanner, B. I. Identification of a lithium interaction site in the γ-aminobutyric acid (GABA) transporter GAT-1. *J. Biol. Chem.* **281,** 22092–22099 (2006).
12. Dodd, J. R. & Christie, D. L. Selective amino acid substitutions convert the creatine transporter to a γ-aminobutyric acid transporter. *J. Biol. Chem.* **282,** 15528–15533 (2007).
13. Vandenberg, R. J., Shaddick, K. & Ju, P. Molecular basis for substrate discrimination by glycine transporters. *J. Biol. Chem.* **282,** 14447–14453 (2007).
14. Radian, R., Bendahan, A. & Kanner, B. I. Purification and identification of the functional sodium- and chloride-coupled γ-aminobutyric acid transport glycoprotein from rat brain. *J. Biol. Chem.* **261,** 15437–15441 (1986).
15. Guastella, J. *et al.* Cloning and expression of a rat brain GABA transporter. *Science* **249,** 1303–1306 (1990).
16. Accardi, A. & Miller, C. Secondary active transport mediated by a prokaryotic homologue of ClC Cl^- channels. *Nature* **427,** 803–807 (2004).
17. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415,** 287–294 (2002).
18. Dutzler, R., Campbell, E. B. & MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **300,** 108–112 (2003).
19. Keynan, S., Suh, Y. J., Kanner, B. I. & Rudnick, G. Expression of a cloned γ-aminobutyric acid transporter in mammalian cells. *Biochemistry* **31,** 1974–1979 (1992).
20. Bennett, E. R., Su, H. & Kanner, B. I. Mutation of arginine 44 of GAT-1, a (Na^+ + Cl^-)-coupled γ-aminobutyric acid transporter from rat brain, impairs net flux but not exchange. *J. Biol. Chem.* **275,** 34106–34113 (2000).
21. Loo, D. D., Eskandari, S., Boorer, K. J., Sarkar, H. K. & Wright, E. M. Role of Cl^- in electrogenic Na^+-coupled cotransporters GAT1 and SGLT1. *J. Biol. Chem.* **275,** 37414–37422 (2000).
22. Lu, C. C. & Hilgemann, D. W. GAT1 (GABA:Na^+:Cl^-) cotransport function. Steady state studies in giant *Xenopus* oocyte membrane patches. *J. Gen. Physiol.* **114,** 429–444 (1999).
23. Ryan, R. M. & Mindell, J. A. The uncoupled chloride conductance of a bacterial glutamate transporter homolog. *Nature Struct. Mol. Biol.* **14,** 365–371 (2007).
24. Mari, S. A. *et al.* Role of the conserved glutamine 291 in the rat γ-aminobutyric acid transporter rGAT-1. *Cell. Mol. Life Sci.* **63,** 100–111 (2006).
25. Mager, S. *et al.* Ion binding and permeation at the GABA transporter GAT1. *J. Neurosci.* **16,** 5405–5414 (1996).
26. Quick, M. & Javitch, J. A. Monitoring the function of membrane transport proteins in detergent-solubilized form. *Proc. Natl Acad. Sci. USA* **104,** 3603–3608 (2007).
27. Forrest, L. R., Tavoulari, S., Zhang, Y.-W., Rudnick, G. & Honig, B. Identification of a chloride ion binding site in Na^+/Cl^- dependent transporters. *Proc. Natl Acad. Sci. USA* **104,** 12761–12766 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a grant from the Israel Science Foundation (to B.I.K.) and NIH grants (to J.A.J.). We thank S. Qu for the generation of the GAT-1 S331G mutant; M. Sonders for making the LeuT E290S and Tyt1 A263S mutants; and L. Chung for cell culture and membrane preparations.

Author Contributions E.Z., A.B. and B.I.K. did the experimental work for the eukaryotic transporters and M.Q. and Y.Z. did the experiments on the bacterial transporters. E.Z., M.Q., J.A.J. and B.I.K. did project planning and manuscript writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.I.K. (kannerb@cc.huji.ac.il).

METHODS

Mutagenesis. Single-stranded, uracil-containing DNA served as template for site-directed mutagenesis of the mammalian transporter genes using appropriate mutagenic primers (usually containing a diagnostic restriction site) followed by second-strand synthesis and ligation[20]. Site-directed mutagenesis of the bacterial transporter genes was performed by PCR with the respective expression vector[10,26] as template and appropriate mutagenic primers. The mutations in the mammalian transporters were identified by analytical restriction analysis. Mutagenic fragments were subcloned back into the wild-type allele using pairs of unique restriction enzymes that cut the cDNA at both sides of the mutation. The constructs were sequenced in both directions between these two restriction sites.

Transport. Expression of the mammalian wild-type and mutant transporters was done as follows: HeLa cells plated on 24-well plates were infected with recombinant vaccinia/T7 virus vTF and transfected with cDNA (pBluescript SK with the wild-type or mutant transporter inserted downstream to the T7 promoter) using the transfection reagent DOTAP as described[19]. Whole-cell uptake of [^{3}H]GABA was assayed 18–20 h after transfection. The wells were washed twice with transport solution (room temperature) containing 150 mM NaCl, 5 mM KPi, pH 7.4, 0.5 mM $MgSO_4$ and 0.3 mM $CaCl_2$. Each well was then incubated with 200 µl of the same solution supplemented with 0.4 µCi of [^{3}H]GABA (Amersham, 94 Ci mmol^{-1}), unless indicated otherwise in the figure legends. Transport was terminated, usually after 10 min, by aspirating the medium and washing twice with 1 ml of transport solution. Cells were lysed with 1% of SDS (0.5 ml well^{-1}) and the radioactivity was measured by liquid scintillation counting. For transport in the absence of chloride, done side by side with that in the presence of chloride, the NaCl and $CaCl_2$ from the transport solution were replaced with the same concentration of the respective gluconate salts. Transport experiments in intact HeLa cells expressing GAT4 and DAT were done as with those expressing GAT-1, except that [^{3}H]dopamine was used with DAT. Transport was usually carried out for 10 min at room temperature (unless indicated otherwise in the figure legends) and the assay was terminated by washing the wells twice with ice-cold NaCl transport solution. Cells were lysed with 1% SDS, and radioactivity was measured by liquid scintillation counting. Uptake of [^{14}C]tyrosine by whole *E. coli* MQ614 expressing wild-type Tyt1 or the given mutants was performed as described[10]. In the experiments shown in Fig. 3f, uptake was assayed for 1 min in the presence of 50 mM Tris/Mes, pH 8.5, 10 mM NaCl or 10 mM Na-gluconate, and 0.5 µM (final concentration) L-[^{14}C(U)]tyrosine (Moravek, 391 Ci mol^{-1}). The Tyt1-specific uptake activity was determined by subtracting the nonspecific background (0-s time point) from the 1-min values. Experiments were done at least three times.

Reconstitution. Reconstitution of the transporters into liposomes using spin columns was done as described[20], and for each transport reaction 10 µl of reconstituted proteoliposomes was added to 360 µl transport solution. For net flux the liposomes contained 120 mM KPi, pH 7.4, or 150 mM KMES, pH 6.0. The external medium consisted of 75 mM NaCl plus either 60 mM NaPi, pH 7.4, or 75 mM NaMES, pH 6.0, supplemented with 2.8 µM valinomycin and 1 µCi of [^{3}H]GABA (94 Ci mmol^{-1}). In those cases where 'in' or 'out' media of different compositions were used, this is indicated. For exchange, the internal medium of the liposomes contained 10 mM GABA, 75 mM NaCl and 60 mM NaPi, pH 7.4 or 75 mM NaMES, pH 6.0. To remove the external unlabelled GABA, the liposomes were passed over a second spin column of identical composition but lacking GABA. The control liposomes (without entrapped GABA) were also passed over a second spin column. The external medium contained 75 mM NaCl plus either 60 mM NaPi, pH 7.4, or 75 mM NaMES, pH 6.0, and 2 µCi of [^{3}H]GABA unless stated otherwise in the figure legends. For chloride-free conditions, gluconate served as substituting anion, using the sodium or potassium salt as appropriate. Transport reactions were terminated by filtration, and radioactivity, retained after washing the filters, was done by liquid scintillation counting as described[14]. Experiments were done at least three times. Although there were differences in absolute activity between reconstitution experiments done on different days, qualitatively the experiments were highly reproducible. Therefore, individual typical experiments are shown in Fig. 2 and in Supplementary Figs 2 and 3. In individual experiments good reproducibility of the data was obtained as demonstrated by several duplicate time points shown.

Substrate binding. Leucine binding by LeuT was performed using a scintillation proximity-based assay as described[26]. Before solubilization, membrane vesicles were thoroughly washed in 200 mM Tris/MES, pH 7.5/1 mM TCEP to eliminate cationic (that is, Na^+) and anionic (that is, Cl^-) contaminations. Purification of each recombinant protein was performed by decahistidine-mediated immobilized Ni^{2+}-affinity chromatography, and the purified proteins were desalted and assayed for subsequent binding experiments as described[26]. Binding studies were performed in 50 mM Tris/MES, pH 7.5/20% glycerol/0.1% (w/v) *n*-dodecyl-β-D-maltopyranoside/1 mM TCEP and either 150 mM NaCl or 150 mM Na-gluconate. A total of 350 ng of purified and desalted wild-type LeuT or E290S mutant protein was used for each individual binding assay. Owing to the lower affinity of E290S for leucine, 100 nM and 1 µM L-[^{3}H]leucine (84 Ci mmol^{-1}) were used for wild-type LeuT and the E290S mutant, respectively.

Cell surface biotinylation. Labelling of wild-type and mutant transporters at the cell surface, using Sulfo-NHS-SS-Biotin (Pierce), quenching the reaction, cell lysis and isolation of the biotinylated proteins by streptavidin-agarose beads (Pierce) was done as described[20]. After SDS–PAGE (10% gel) and transfer to nitrocellulose, the GAT-1 protein was detected with an affinity-purified antibody, directed against an epitope from the cytoplasmic C-terminal tail of GAT-1, at a 1:1,000 dilution, with horseradish peroxidase-conjugated secondary antibody at a 1:40,000 dilution, and with ECL. One per cent of goat serum was present in all antibody, blocking and washing solutions to minimize the appearance of nonspecific bands. Scanning and quantification of bands on the developed films was done using a Fluor-S MultiImager scanner with MultiAnalyst software, both from Bio-Rad.

Homology modelling. This was done using the Tripos Sybyl7.0 computational software as described[11]. The details are given in the legend to Fig. 4.

LETTERS

UTX and JMJD3 are histone H3K27 demethylases involved in *HOX* gene regulation and development

Karl Agger[1,2]*, Paul A. C. Cloos[1,2]*, Jesper Christensen[1,2]*, Diego Pasini[1,2], Simon Rose[1], Juri Rappsilber[3], Irina Issaeva[4], Eli Canaani[4], Anna Elisabetta Salcini[1] & Kristian Helin[1,2]

The trithorax and the polycomb group proteins are chromatin modifiers, which play a key role in the epigenetic regulation of development, differentiation and maintenance of cell fates[1–3]. The polycomb repressive complex 2 (PRC2) mediates transcriptional repression by catalysing the di- and tri-methylation of Lys 27 on histone H3 (H3K27me2/me3)[3]. Owing to the essential role of the PRC2 complex in repressing a large number of genes involved in somatic processes, the H3K27me3 mark is associated with the unique epigenetic state of stem cells[4–7]. The rapid decrease of the H3K27me3 mark during specific stages of embryogenesis and stem-cell differentiation indicates that histone demethylases specific for H3K27me3 may exist. Here we show that the human JmjC-domain-containing proteins UTX and JMJD3 demethylate tri-methylated Lys 27 on histone H3. Furthermore, we demonstrate that ectopic expression of JMJD3 leads to a strong decrease of H3K27me3 levels and causes delocalization of polycomb proteins *in vivo*. Consistent with the strong decrease in H3K27me3 levels associated with *HOX* genes during differentiation, we show that UTX directly binds to the *HOXB1* locus and is required for its activation. Finally mutation of F18E9.5, a *Caenorhabditis elegans* JMJD3 orthologue, or inhibition of its expression, results in abnormal gonad development. Taken together, these results suggest that H3K27me3 demethylation regulated by UTX/JMJD3 proteins is essential for proper development. Moreover, the recent demonstration that UTX associates with the H3K4me3 histone methyltransferase MLL2 (ref. 8) supports a model in which the coordinated removal of repressive marks, polycomb group displacement, and deposition of activating marks are important for the stringent regulation of transcription during cellular differentiation.

The recent identification of LSD1 and JmjC-domain-containing histone lysine demethylases[9,10] suggests that histone methylation marks are more dynamic than previously anticipated. We and others have previously noticed that the phylogenetic subgrouping of the jumonji protein family, based on the homologies in the JmjC domains, correlates with their enzymatic specificities[9–12]. In an attempt to identify potential H3K27 demethylases amongst the JmjC-domain-containing proteins, we noticed that JMJD3 a member of the phylogenetic subcluster consisting of UTX, UTY and JMJD3 (Fig. 1a) is highly expressed in embryonic stem cells[4]. The ubiquitously transcribed tetratricopeptide repeat X (*UTX*) gene is localized on the X chromosome and escapes X-chromosome inactivation[13]. UTY (ubiquitously transcribed tetratricopeptide repeat Y) seems to represent the male counterpart to UTX and is encoded by the male Y chromosome[13]. Both proteins are highly homologous and share the same domain structure, including TPR protein interaction domains. JMJD3 shows high homology and structural relationship to UTX/

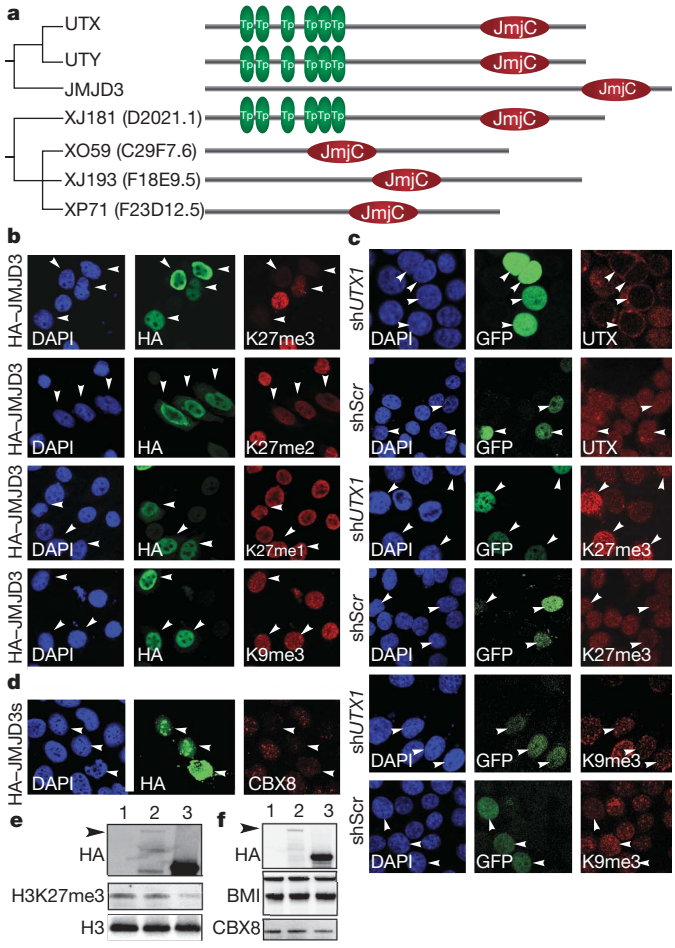


Figure 1 | UTX and JMJD3 are H3K27me3 demethylases. a, Phylogenetic cluster containing UTX, UTY, JMJD3 and homologous proteins of *C. elegans*. Tp, tetratricopeptide repeat; JmjC, jumonji C domain. **b**, HeLa cells transfected with *pCMV-HA–JMJD3*; white arrowheads indicate transfected cells. **c**, HeLa cells co-transfected with shRNA to *UTX* (sh*UTX1*) or a scrambled control (shScr) and *pCMV-H2B–GFP* using a ratio of 10:1. **d**, U2OS cells transfected with *pCMV-HA–JMJD3s*, pre-extracted and co-stained for HA and CBX8. **e**, Immunoblot using lysates from HeLa cells transfected with control (lane 1), *pCMV-HA–JMJD3* (lane 2) or *pCMV-HA–JMJD3s* (lane 3). An arrowhead indicates the full-length JMJD3 protein. **f**, Immunoblot with antibodies to HA (JMJD3), CBX8 and BMI1 using lysates from U2OS cells transfected as in panel **d**.

[1]Biotech Research and Innovation Centre (BRIC), and [2]Centre for Epigenetics, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. [3]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK. [4]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel.
*These authors contributed equally to this work.

UTY, especially in the JmjC domain, but lacks the TPR domains. In *C. elegans* this subfamily is represented by four genes, respectively encoding one protein related to UTX/UTY and three related to JMJD3 (Fig. 1a). Taken together with the important role of the H3K27me3 mark in pluripotent embryonic stem cells, and the down-regulation of this mark during differentiation into somatic cells[7], we speculated that the UTX-family could have H3K27me3 demethylase activity.

To test the activity of the UTX and JMJD3 proteins, we cloned the full-length open reading frame of the genes into a mammalian expression vector. As shown in Fig. 1b, ectopic expression of JMJD3 markedly reduced the levels of H3K27me3 and to a lesser degree H3K27me2, whereas we did not observe any decrease of H3K27me1. Furthermore no decrease was observed on a number of other methylated lysines, including H3K9me3, H3K4me3/me2, H3K36me3 and H1K26me3 (Supplementary Fig. 1), suggesting that JMJD3 specifically demethylates H3K27me3/me2. In agreement with this, ectopic expression of haemagglutinin-tagged JMJD3 (HA–JMJD3) as well as a construct expressing the short carboxy-terminal part of JMJD3 encompassing the JmjC domain (amino acids 1027–1685, HA–JMJD3s) led to globally reduced H3K27me3 levels (Fig. 1e). To test for the specificity of the reduction in H3K27me3 levels we generated a HA–JMJD3s mutant (H1392T/E1394G), in which two of the three amino acids in the JmjC domain believed to be involved in Fe(II) binding were mutated. Indeed, when this mutant was ectopically expressed in HeLa cells, no reduction in H3K27me3 methylation levels was observed (Supplementary Fig. 2), demonstrating that the demethylating activity of JMJD3 is dependent on a catalytically active JmjC domain.

In contrast to JMJD3, ectopic expression of UTX did not result in a significant reduction of H3K27me3 levels (data not shown). This result could be due to limiting amounts of co-factors required for UTX recruitment to chromatin and/or high levels of endogenous UTX. Thus, to understand if endogenous UTX is involved in the demethylation of H3K27me3, we generated constructs expressing short hairpin (sh)RNAs specifically targeting *UTX*. The transfection of these constructs significantly depleted both *UTX* messenger RNA (see later) and protein (Supplementary Fig. 3). We co-transfected the shRNA construct with an histone H2B (HIST1H2BJ)–GFP (green fluorescent protein) expression plasmid enabling identification of the transfected cells by fluorescence. As shown in Fig. 1c, the expression of the shRNA complementing UTX resulted in a significant decrease in UTX expression, and a concomitant increase of H3K27me3 levels, whereas no change in H3K9me3 levels was observed. Moreover, we observed a slight increase in H3K27me3 levels in both HeLa and 293 cells, as assessed by immunoblotting (Supplementary Fig. 3). These experiments suggest that UTX is important for maintaining the steady-state levels of H3K27me3 in proliferating cells.

Polycomb-mediated repression requires the presence of H3K27me3/me2, which in part functions as a recruitment template for the chromodomain-containing proteins of the PRC1 complexes[2]. As shown in Fig. 1d, ectopic expression of JMJD3 and JMJD3s (amino acids 1027–1685, containing the JmjC domain) resulted in the loss of the PRC1 member and chromodomain-containing protein, CBX8 (ref. 14) from polycomb bodies without affecting the expression of CBX8 and BMI1 (Fig. 1f and Supplementary Fig. 3). Taken together, these results suggest that JMJD3 and UTX may antagonize the formation and maintenance of polycomb-group-mediated repression. To determine the enzymatic activity of the UTX/UTY/JMJD3 group, we affinity-purified full-length human UTX from insect cells (Fig. 2a). Recombinant UTX was further purified by size-exclusion chromatography (Supplementary Fig. 4). As shown in Fig. 2b, UTX efficiently reduced the levels of H3K27me3 and H3K27me2, whereas no reduction was observed for any other of the tested methylated lysine marks. These results suggest that UTX is a specific demethylase for H3K27me3 and H3K27me2 *in vitro*.

To demonstrate formally that UTX demethylates H3K27me3 and H3K27me2, synthetic H3K27me3/me2/me1 peptides were incubated with buffer alone or in the presence of UTX using an enzyme to substrate ratio of approximately 1:15 (Fig. 2c). Mass spectrometry analysis revealed that about 50% of the H3K27me3 peptide was converted to H3K27me2 and H3K27me1, and a more moderate conversion of the other peptides. These results indicate that the order of substrate preference is H3K27me3 > me2 > me1. Although no conversion of H3K27me1 was observed for histones *in vivo* and *in vitro*, the peptide data suggest that under specific reaction conditions UTX might also remove H3K27me1.

Next, we tested if JMJD3 could demethylate H3K27me3 *in vitro*, and, as shown Fig. 2d, recombinant JMJD3s (Fig. 2a) indeed demethylated H3K27me3, whereas virtually no reactivity was observed for H3K27me2/me1. In addition, no changes were observed for a panel of other histone H3 methylation marks (Fig. 2d), confirming the specificity for H3K27me3.

Interestingly UTX was recently identified as a component of a multiprotein complex containing a trithorax-related protein MLL2/ALR[8]. Here we have shown that UTX, by demethylating H3K27me3/me2, can work antagonistically to the polycomb group proteins, and therefore has features of a trithorax-group protein. Trithorax and polycomb group proteins are well-known regulators of *HOX* genes in both *Drosophila* and mammals[1–3]. Previous results have shown that expression of *HOX* genes is tightly regulated during retinoic acid-induced differentiation of human embryonic teratocarcinoma NT2/D1 cells, and that this is associated with a fast modulation of H3K27me3 levels at the promoters of these genes[5,15]. We therefore

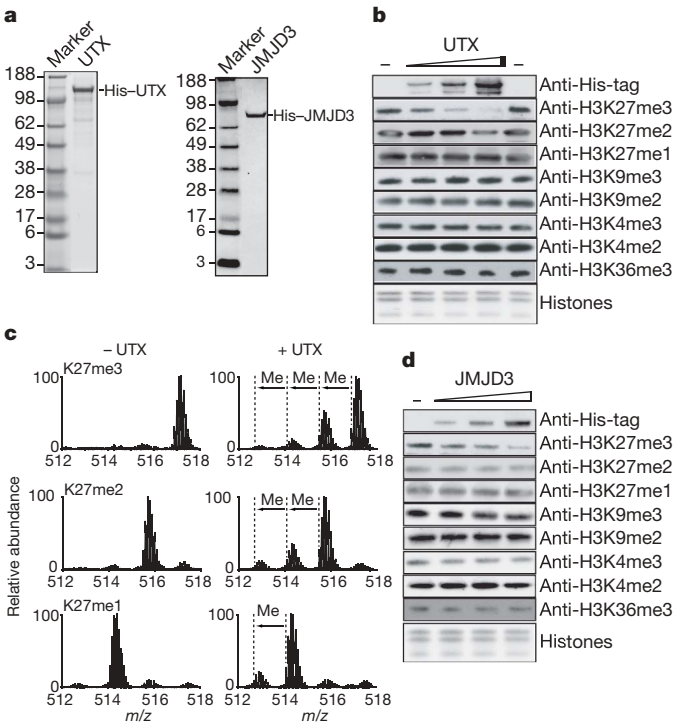


Figure 2 | UTX and JMJD3 demethylate H3K27me3/me2 *in vitro*. **a**, Recombinant UTX and JMJD3, separated by SDS–PAGE and stained by Coomassie blue. **b**, Demethylation activity of recombinant UTX using histones as the substrate and analysed by immunoblotting. Ponceau staining for histones was used as a loading control. **c**, H3K27me3/me2/me1 peptides (4 μg) were incubated with or without recombinant UTX (8 μg) and analysed by mass spectrometry. The masses of tri-, di-, mono-, and un-methylated H3K27 peptides were 5,158.916, 5,144.900, 5,130.885 and 5,116.869 Da, respectively. Shown is the charge state +10. A shift in mass equivalent to one methyl group is indicated as 'Me'. **d**, Demethylation activity of recombinant JMJD3 using histones as the substrate and assayed by immunoblotting. DAPI, 4,6-diamidino-2-phenylindole staining.

tested if UTX could be involved in the retinoic-acid-induced reduction of H3K27me3 levels at *HOX* genes in NT2/D1 cells. To do this, we treated NT2/D1 cells with retinoic acid for 0 h and 24 h and measured the expression of *HOXB1* mRNA by reverse transcription real-time quantitative PCR (qPCR). In agreement with previously published results, retinoic acid treatment led to a strong increase of *HOXB1* mRNA levels (Fig. 3b). To test a potential role for UTX in the regulation of *HOXB1* expression, we performed chromatin immunoprecipitation (ChIP) assays on cells harvested at 0 h and 24 h after treatment with retinoic acid. As shown in Fig. 3b, UTX was bound at the *HOXB1* promoter in the absence of retinoic acid, and retinoic acid treatment led to a twofold increase in UTX binding within 24 h. Retinoic acid treatment led to a concomitant decrease in H3K27me3 levels as well as an increase in H3K4me3 levels (Fig. 3b). Similar results were obtained in mouse embryonic stem cells, which showed loss of H3K27me3 methylation and dissociation of polycomb group proteins (Supplementary Fig. 5). In addition, retinoic acid treatment also led to increased binding of MLL2 to the *Hoxb1* promoter (Supplementary Fig. 5). Taken together with the previous finding that UTX is found in a complex with MLL2, these results indicate that retinoic-acid-induced differentiation results in the recruitment and/or activation of a complex containing H3K27me3 demethylase activity and H3K4me3 histone methyltransferase activity that mediates transcriptional activation of the *HOX* promoters. Consistent with this model, the inhibition of UTX expression by shRNA (Fig. 3c) led to a strong decrease in *HOXB1* mRNA levels both before and after retinoic acid treatment (Fig. 3c), as well as an increased level of H3K27me3 at the *HOXB1* promoter (Fig. 3d). No significant change in the amount of H3 bound at the promoter was observed (Fig. 3d).

C. elegans has four protein sequences with the assigned names XJ193, XO59, XP71 and XJ118, (gene sequence identifiers *F18E9.5*, *C29F7.6*, *F23D12.5* and *D2021.1*, respectively) with strong homology to the UTX/UTY/JMJD3 family members. Although the predicted D2021.1 protein contains both tetratricopeptide repeats (Tpr) and a JmjC domain, suggesting that it is a UTX homologue, the other three members of the family only contain a JmjC domain, and therefore are more similar to the JMJD3 protein (Fig. 1a). To identify the *in vivo* role of this class of demethylases, we characterized the function of a member of the *C. elegans* JMJD3 family, *F18E9.5*. First, we showed that recombinant full-length XJ193 (Fig. 4b) specifically demethylates H3K27me3 (Fig. 4c). Second, we investigated the functional role of the *F18E9.5* gene in *C. elegans* development. To do this we analysed two strains (*VC936* and *VC912*) carrying mutant alleles of *F18E9.5*. *VC936* and *VC912* strains (with the corresponding alleles *gk384* and *gk387*) contain large deletions in the 5′-end of the *F18E9.5* sequence (Fig. 4a), leading to a frame-shift that results in a disrupted open reading frame. Both strains showed a common gonadal phenotype particularly evident at high temperature (25 °C). As shown in Fig. 4d,

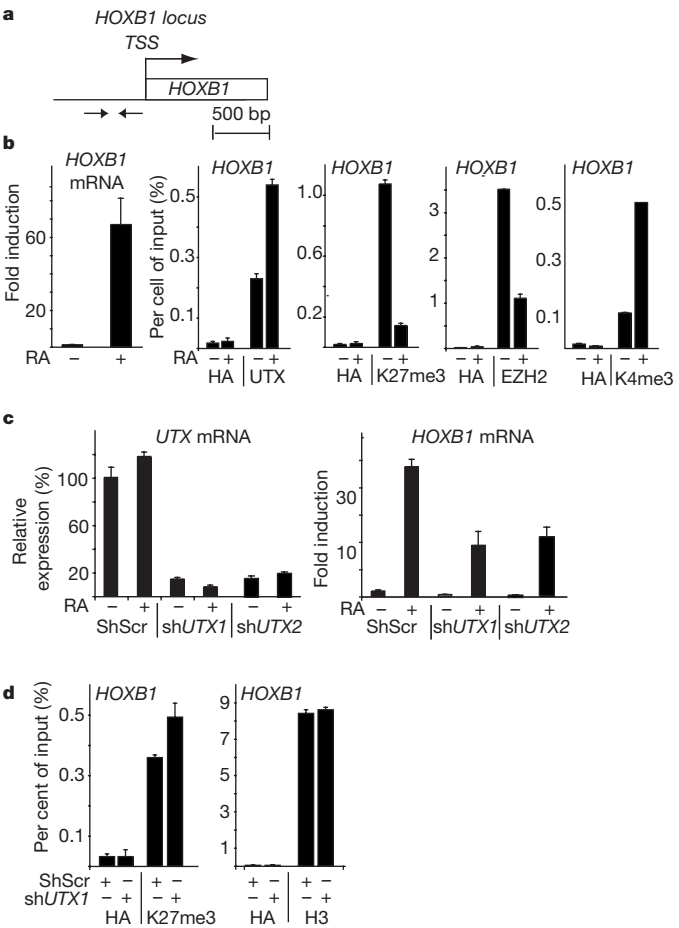


Figure 3 | UTX is required for transcriptional activation of *HOXB1* during differentiation. **a**, Human *HOXB1* locus and primers used for ChIP experiments. **b**, *HOXB1* mRNA in NT2/D1 cells after 24 h of treatment with 1 μM trans retinoic acid (RA) as measured by qPCR (left panel). Other panels, ChIP using NT2/D1 cells, untreated or treated with RA. Enrichment of the *HOXB1* promoter was measured by qPCR. **c**, Left panel, knockdown efficiency of sh*UTX1* and sh*UTX2* in NT2/D1 cells assayed by qPCR. Right panel, induction of *HOXB1* mRNA after RA treatment. **d**, ChIP using NT2/D1 cells transfected with sh*UTX1* or shScr, and treated with RA. Error bars provide s.d., $n = 3$.

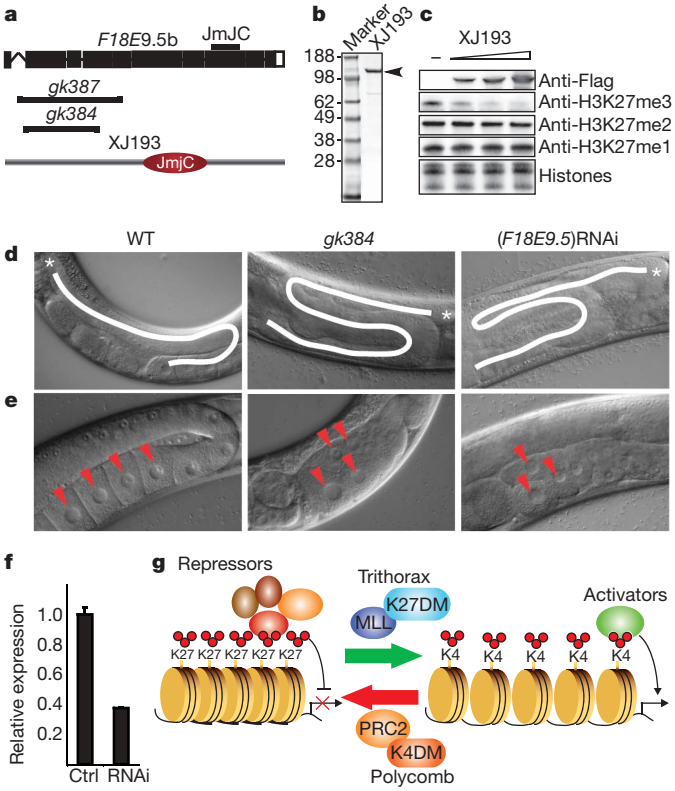


Figure 4 | The JMJD3 homologue, XJ193 is required for gonadal development. **a**, Genomic organization of *F18E9.5*, encoding for XJ193. **b**, SDS–PAGE showing recombinant XJ193. **c**, Demethylation activity of recombinant XJ193 using histones as the substrate. **d**, Defects in gonad migration in *gk384* and (*F18E9.5*)RNAi animals. Asterisks indicate the distal end of the gonad. Lines indicate distal tip cell path finding. **e**, Accumulation of oocytes in *gk384* and (*F18E9.5*)RNAi animals in the proximal gonad arm (arrowheads). **f**, Relative expression of *F18E9.5* mRNA in control and RNAi-treated animals. Error bars indicate s.d., $n = 3$. **g**, Model for the involvement of H3K27me3 demethylases in the transcriptional regulation of developmental genes. K27DM, Lys 27 demethylase; K4DM, Lys 4 demethylase. Small red circles, methyl groups.

gonad migration was aberrant in a high fraction (in *gk384*: 128/200, 64%) of adult mutant hermaphrodites with aberrant turns and positioning of the distal ends. Furthermore, for a lower fraction of the mutant adult animals (in *gk384:* 59/200, 30%), the proximal end of the gonad was enlarged with a disorganized accumulation of nuclei in the diakinesis phase (Fig. 4e). Importantly, both phenotypes were also present, albeit at a lower frequency (gonad migration, 12/110, 10%; oocyte accumulation 10/110, 9%), in (*F18E9.5*)RNA interference (RNAi)-treated animals (Fig. 4d, e). This result supports the role of the *F18E9.5* gene in gonadal development and organization. Moreover, the reason for the higher frequency of aberrant gonadal development in the mutant worms is most probably because the mutants analysed are null alleles, whereas the RNAi treatment only led to a 2.5-fold reduction in the expression of *F18E9.5* mRNA (Fig. 4f). Consistent with the gonadal defects, the brood size of the mutant animals was significantly reduced compared with wild type (*gk384*, 110 ± 19.5; N2 (wild type): 228 ± 28 at 25 °C; $P < 0.0012$ (numbers indicate means $\pm$ s.d.)).

Tri- and di-methylation of H3K27 is believed to be the key mechanism by which the PRC2 complex regulates a number of cellular processes such as development, differentiation and proliferation. The H3K27me3/me2 epigenetic marks are, in general, associated with transcriptional repression, and, consistent with the functional role of the PRC2 complex, the target genes of the complex are a blueprint of genes controlling cell fate decisions[1–3]. Even though the H3K27me3/me2 marks most probably are required as a long-term 'memory' system to maintain transcriptional silencing, the H3K27me3/me2 marks are rapidly turned over during cellular differentiation and at the *INK4A-ARF* locus during senescence[7,16]. The demonstration that the UTX and JMJD3 proteins are lysine demethylases, which can specifically demethylate H3K27me3 and H3K27me2, indicate that they could be key regulators of important cellular decisions. In agreement with this, we have shown that the JMJD3 homologue XJ193 is essential for proper gonadal development in *C. elegans*, and that UTX participates in the regulation of *HOX* genes during differentiation of teratocarcinoma and embryonic stem cells.

Interestingly, UTX has previously been shown to be part of the MLL2 complex. The finding of a histone demethylase for H3K27me3 in a complex with a H3K4me3 methyltransferase (Fig. 4g) may indicate an elegant mechanism by which the removal of a repressive mark and the simultaneous deposition of an active mark can lead to the activation of a target gene. Whether this is a general mechanism for gene regulation remains to be established; however, it is tempting to speculate that lysine methyl transferases involved in transcriptional repression (such as the PRC2 complex or SUV39H1) could more effectively regulate transcription by binding to H3K4me3 demethylases (Fig. 4g).

METHODS SUMMARY

Cloning procedures. The open-reading frame (ORF) of human *UTX* was PCR amplified from a human fetal brain complementary DNA library and cloned into the pEntr3C entry gateway vector (Invitrogen). The ORF of *JMJD3* was PCR amplified from KIAA0346 and cloned into pEntr3C. The PCR-generated ORFs were sequence-verified. The expression constructs pCMV-HA, pCMV-HA-NLS, baculovirus virus transfer vector pVL-Flag-His and pACHlt-a, were generated using the Gateway system (Invitrogen).

Generation of antibodies to UTX. Polyclonal antibodies were generated by immunizing rabbits with affinity-purified GST–UTX (amino acids 453–753) expressed in *E. coli*. The antisera were absorbed on glutathione *S*-transferase (GST)-coupled cyanogen-bromide-activated Sepharose (Amersham) and, subsequently, affinity-purified using GST–UTX-coupled Sepharose. Polyclonal anti MLL2 (ALR) was described[8].

***C. elegans* strains and methods.** Maintenance, culturing and genetic manipulations of *C. elegans* strains were carried out according to standard procedures[17] and conducted at 15–18 °C. However, the mutant phenotypes were observed at 25 °C.

Immunofluorescence analysis. Immunofluorescence analysis was performed as described[11,12].

Demethylation assay. Demethylation assays were performed as described[11,12].

Expression analysis and ChIP assays. Real-time quantitative PCR analysis and ChIP assays were performed with the indicated antibodies, essentially as described[18].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 July; accepted 3 August 2007.
Published online 22 August 2007.

1. Ringrose, L. & Paro, R. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* **134,** 223–232 (2007).
2. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128,** 735–745 (2007).
3. Schwartz, Y. B. & Pirrotta, V. Polycomb silencing mechanisms and the management of genomic programmes. *Nature Rev. Genet.* **8,** 9–22 (2007).
4. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
5. Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **20,** 1123–1136 (2006).
6. Lee, T. I. *et al.* Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* **125,** 301–313 (2006).
7. Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128,** 747–762 (2007).
8. Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* **27,** 1889–1903 (2007).
9. Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. *Nature Rev. Mol. Cell Biol.* **8,** 307–318 (2007).
10. Shi, Y. & Whetstine, J. R. Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell* **25,** 1–14 (2007).
11. Christensen, J. *et al.* RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell* **128,** 1063–1076 (2007).
12. Cloos, P. A. *et al.* The putative oncogene *GASC1* demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* **442,** 307–311 (2006).
13. Greenfield, A. *et al.* The *UTX* gene escapes X inactivation in mice and humans. *Hum. Mol. Genet.* **7,** 737–742 (1998).
14. Dietrich, N. *et al.* Bypass of senescence by the polycomb group protein CBX8 through direct binding to the *INK4A-ARF* locus. *EMBO J.* **26,** 1637–1648 (2007).
15. Lee, V. M. & Andrews, P. W. Differentiation of NTERA-2 clonal human embryonal carcinoma cells into neurons involves the induction of all three neurofilament proteins. *J. Neurosci.* **6,** 514–521 (1986).
16. Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the *INK4A-ARF* locus and are disassociated in senescent cells. *Genes Dev.* **21,** 525–530 (2007).
17. Brenner, S. The genetics of *Caenorhabditis elegans. Genetics* **77,** 71–94 (1974).
18. Bracken, A. P. *et al. EZH2* is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22,** 5323–5335 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to U. Toftegaard and R. Christensen for expert technical assistance. We thank T. Stiernagle for her continuing assistance with strain requests. We also thank the *C. elegans* reverse genetic core facility, Vancouver, Canada, for the production of the mutant alleles. We thank members of the Helin and Rappsilber laboratories for technical advice and support. P.A.C.C. was supported by the Benzon Foundation, and D.P. by a post-doctoral fellowship from the Danish Medical Research Foundation. A Marie Curie Excellence Grant from the European Commission supported J.R. The work in the Helin laboratory was supported by grants from the Danish Cancer Society, the Novo Nordisk Foundation, the Danish Medical Research Council, the Danish Natural Science Research Council, the Danish National Research Foundation, and the International Association for Cancer Research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.H. (kristian.helin@bric.dk).

METHODS

Cloning procedures. A plasmid, pF1KSDA0346, encoding the ORF of human JMJD3 (KIAA0346) was obtained from the cDNA Bank Section Kazusa DNA Research Institute, Japan. Sequence verification of the plasmid revealed two coding mutations in the JmjC domain of the gene that are not listed in the database and that inactivate the enzymatic activity of the protein. The JMJD3 ORF was subsequently transferred into pEntr3C and the mutations were corrected by substituting the JmjC domain with sequences obtained by PCR-amplification from the cDNA library described above. *F18E9.5* was cloned by reverse transcription PCR cloning using *C. elegans* RNA as a template. The amplified fragment was cloned into pCR8/GW Gateway entry vector. Full-length assembled UTX, JMJD3 and XJ193 were verified by sequencing. 5′ deletions of *UTX* and *JMJD3* were generated by PCR and fragments were cloned into pCR8/GW. To generate expression vectors, the entry clones were transferred into Gateway-compatible pCMV-HA, and Gateway compatible baculovirus virus transfer vector pVL-Flag-His and pACHlt-a, which encodes amino-terminal Flag- and hexa-histidine tags. Recombinant baculoviruses were generated by co-transfection of baculovirus transfer vector encoding UTX, JMJD3 and XJ193, and *Bsu*36I-linearized Bakpak6 baculovirus DNA. Histidine-tagged proteins were expressed in Trichoplusia ni, High Five, or Sf9 cells and purified by cobalt–Sepharose affinity chromatography. Selected fractions were further purified by size-exclusion chromatography as described previously[12].

shRNA procedures. shRNA constructs targeting two different sequences, sh*UTX1* (CTATGAATCTCTAATCTT) and sh*UTX2* (TAGCTTTGGTTGACTGTA), were used in these studies. The plasmids were transfected into NT2/D1 cells using Fugene 6. After transfection the cells were selected for 48 h with puromycin ($2\,\mu g\,ml^{-1}$).

Demethylation assay. Bulk histones were incubated with purified His–UTX and His–JMJD3 for 60 min at 37 °C in demethylation buffer (50 mM Hepes-KOH, pH 7.7, 25 mM KCl, 1 mM $MgCl_2$, 1 mM ketoglutarate, 50 μM $FeSO_4$, 2 mM ascorbic acid). Reaction mixtures were analysed by either western blotting using specific antibodies or mass spectrometry. The following antibodies were used: anti-H3K9me3 (Abcam ab8898), anti-H3K9me2 (Abcam ab1220), anti-H3K27me3 (Upstate 07-449), anti-H3K27me2 (Abcam 24684), H3K27me (Abcam 24684), anti-H3K4me3, (Abcam ab8580 and Abcam ab6000), anti-H3K4me2 (Abcam ab7766), anti H3K36me3 (Abcam ab9050), anti-H3K36me2 (Upstate 07-369), anti-H3 (Abcam Ab1791-100), anti-His (Abcam ab9108), and anti-HA (CRP AFC-101P).

Mass spectrometry analysis. Eight micrograms of recombinant UTX were incubated with 4 μg of H3K27me3, H3K27me2, or H3K27me1 peptide with the sequence ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHR-YC-Ttds-K-Biotin synthesized by Jerini in demethylation buffer in a final volume of 170 μl for 60 min at 37 °C. Aliquots of the reaction mixtures corresponding to 0.2 μg of peptide were desalted on a C18-StageTip[19]. The peptides were then injected into an LC-MS system composed of a PAL autosampler (CTC Analytics), nanoflow HPLC pump (Agilent), and LTQ-Orbitrap mass spectrometer (ThermoElectron). The spray emitter of 75 μm diameter was packed with reversed phase beads to give a 7 cm column. A gradient of acetonitrile with 0.5% acetic acid was run from 0–10% in 75 min at $300\,nl\,min^{-1}$.

Immunofluorescense analysis. HeLa, or U2OS cells were transfected with the indicated expression or shRNA-encoding constructs using Fugene or the $CaPO_4$ method. Post infection (24–48 h) the transfected cells were fixed, permeabilized and stained with the appropriate antibodies and secondary fluorochrome labelled antibodies. The stained cells were analysed by confocal laser fluoroscense microscopy on a Carl Zeis, Axiovert 200M confocal microscope. Cell lysates were analysed by western blotting using the indicated antibodies. The loading was normalized to histone H3. When staining for PRC1 bodies, the coverslips were pre-extracted with 20 mM HEPES, pH 7.2, 0.5% IGEPAL-630, 50 mM NaCl, 3 mM $MgCl_2$ and 300 mM sucrose before fixation. When staining for PRC1 bodies, the coverslips were pre-extracted with 20 mM HEPES, pH 7.2, 0.5% IGEPAL-630, 50 mM NaCl, 3 mM $MgCl_2$ and 300 mM sucrose before fixation. In addition to the generated UTX antibody the following antibodies were used: anti-H3K27me3 (Upstate 07-449), anti-H3K27me2 (Abcam 24684), H3K27me (Abcam 24684), anti-H3K9me3 (Abcam ab8898), anti-H3K4me3, (Abcam ab8580 and Abcam ab6000), anti-H3K4me2 (Abcam ab7766), anti H3K36me3 (Abcam ab9050), anti-H1K26me3 (Abcam 17347), anti-HA (Covance PRB-101c and Covance MMS-101p).

Expression analysis and ChIP assays. ChIP assays were performed and analysed as previously described[18]. NT2 and mouse embryonic stem cells were induced with all-*trans* retionic acid (ATRA), as described in the figure legends. In addition to the generated UTX antibody, rabbit anti-HA (Santa Cruz), rabbit anti-H3K4me3 (Abcam ab8580), rabbit anti-H3K27me3 (Upstate 07-449), UTX (ref. 8), MLL2 (ref. 8) and rabbit anti-H3 (Abcam ab1791) were used for ChIP. Total RNA was purified from NT2/D1 cells using RNAeasy (Qiagen). Total RNA was reverse transcribed using a TaqMan reverse transcription reagents from ABI, according to the manufacturer's instructions. For RNA quantification, reversed-transcribed total RNA was analysed by real-time PCR using SYBR Green PCR Master Mix and an ABI prism 7700 Sequence Detection system. All reactions were analysed in triplicates. Primer sequences are available on request.

***C. elegans* strains and methods.** The Bristol strain (N2) was used as the wild-type strain. *VC936* and *VC912* were both obtained from the CGC centre. The *C. elegans F18E9.5* sequence is located on chromosome X and encodes for two transcripts. The longer isoform (*F18E9.5b*) encompasses 10 exons coding for a predicted protein of 1061 amino acids. The ATG of the gene is located at position 7,749 bp of the F18E9 cosmid (U29614) and a terminator codon TAA at position 11,709 bp. The shorter isoform, lacking the first exon, encodes for a protein of 1020 amino acids and the ATG is located in position 8,144 bp of the F18E9 cosmid. Two alleles of *F18E9.5* were identified in the *C. elegans* reverse-genetic core facility, Vancouver. The *gk384* allele (*VC936* strain) lacks 1,175 bp and the deletion is at position 7,954–9,129 bp of the GenBank entry U29614. The sequence also revealed an insertion of 16 nucleotides, as annotated in worm base. The *gk387* allele (*VC912* strain) lacks 1,600 bp and the deletion is located at position 7,910–9,510 bp of the GenBank entry U29614. Both deletions were confirmed by sequencing the locus of the two alleles. The *gk387* allele was backcrossed with N2 two times, and *gk384* was backcrossed with N2 three times. The *gk384* allele was used for statistical analysis of the phenotypes but both alleles shared the same gonadal phenotypes. Synchronized larval stages were obtained from eggs prepared by hypochlorite treatment, plated on NGM plates with OP50 bacteria as food.

Brood size. Five L4 animals from N2 and *gk384* allele were singularly plated in NGM plates and the number of progeny scored after 4 days at 25 °C.

RNA interference. RNA interference was obtained by feeding, carried out as described previously[20]. A clone from the *C.elegans* RNAi feeding library (J. Ahringer's laboratory), containing the region spanning between 9,451 and 10,611 bp of the GenBank entry U29614, was obtained and IPTG-induced bacteria were seeded in NGM agar plates. Eggs were added onto the plates and cultivated at 25 °C. Control animals were fed with bacteria carrying an empty L4440 vector. The efficiency of mRNA down-regulation was tested by qPCR as previously described[11]. P_0 and F_1 progeny were scored for effects. The gonadal phenotype, evident in the F_1 generation, was scored by observing young adult animals, using DIC optics. The reduction of *F18E9.5* observed by qPCR after RNAi is modest and could explain the low percentage of phenotype observed in the interfered animals.

19. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
20. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112 (2001).

LETTERS

Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4

Takayuki Obita[1], Suraj Saksena[2], Sara Ghazi-Tabatabai[1], David J. Gill[1], Olga Perisic[1], Scott D. Emr[2] & Roger L. Williams[1]

The AAA+ ATPases are essential for various activities such as membrane trafficking, organelle biogenesis, DNA replication, intracellular locomotion, cytoskeletal remodelling, protein folding and proteolysis[1]. The AAA ATPase Vps4, which is central to endosomal traffic to lysosomes[2,3], retroviral budding[4] and cytokinesis[5], dissociates ESCRT complexes (the endosomal sorting complexes required for transport) from membranes[6–15]. Here we show that, of the six ESCRT-III-related subunits in yeast, only Vps2 and Did2 bind the MIT (microtubule interacting and transport) domain of Vps4, and that the carboxy-terminal 30 residues of the subunits are both necessary and sufficient for interaction. We determined the crystal structure of the Vps2 C terminus in a complex with the Vps4 MIT domain, explaining the basis for selective ESCRT-III recognition. MIT helices α2 and α3 recognize a (D/E)xxLxxRLxxL(K/R) motif, and mutations within this motif cause sorting defects in yeast. Our crystal structure of the amino-terminal domain of an archaeal AAA ATPase of unknown function shows that it is closely related to the MIT domain of Vps4. The archaeal ATPase interacts with an archaeal ESCRT-III-like protein even though these organisms have no endomembrane system, suggesting that the Vps4/ESCRT-III partnership is a relic of a function that pre-dates the divergence of eukaryotes and Archaea.

The Vps4 AAA ATPase has a key function in the endocytic pathway for the degradation of membrane proteins, and it is recruited to endosomes by means of an N-terminal MIT domain[7,16]. On binding ATP, Vps4 assembles into an oligomer thought to consist of two stacked hexameric rings[7,17]. Proteins bound to the rings experience mechanical forces coupled to conformational changes that occur during catalysis.

Vps4 function is tightly linked to ESCRTs, the multiprotein complexes required for the sorting of endocytosed transmembrane proteins into intraluminal vesicles (ILVs) of morphologically distinctive endosomes known as multivesicular bodies (MVBs). Inward budding of the endosomal limiting membrane generates ILVs, and subsequent fusion of endosomes with lysosomes leads to the hydrolysis of ILVs and their cargo proteins. Vps4 deletion results in an accumulation of all ESCRTs on endosomes[2,3]. ESCRT-III subunits transiently recruit Vps4 to endosomes.

There are six families of ESCRT-III-like proteins[8,18], namely the Vps2, Vps20, Vps24, Snf7, Did2 and Vps60 subunits in yeast. ESCRT-III subunits have a basic N-terminal four-helical core[19] and an acidic C-terminal region. They can homodimerize and heterodimerize through their N-terminal regions, but they also have a metastable, monomeric, closed form in which the C-terminal region blocks heterodimerization[20,21]. The C-terminal region of some ESCRT-III subunits can also recruit MIT-domain-containing proteins such as Vps4 and the deubiquitinase AMSH[17,21–23]. Several ESCRT-III subunits can interact directly with membranes[18,20,24,25]; combined with their ability to homodimerize and heterodimerize, this means that the ESCRT-III subunits can form lattices with great avidity for membranes. This may account for the requirement of an ATPase to disassemble such lattices efficiently (Supplementary Fig. 1).

Vps2 and Vps24 are required for the recruitment of Vps4 to endosomes[6,18]. Both full-length Vps2 and C-terminal constructs bind directly to the Vps4 MIT domain (Fig. 1a–c). A construct consisting only of residues 183–232 binds the MIT domain with a K_d of 28 μM (Fig. 1d), and we determined the 2.0 Å resolution structure of it in a complex with the MIT domain of Vps4 (Supplementary Table 1).

The MIT domain consists of three antiparallel helices arranged in a slightly asymmetric fashion (Fig. 1e, g). The MIT-interacting region of Vps2 has three short helical segments, αA, αB and αC (Fig. 1e). Helices αA and αB are single-turn helices joined through a turn at Pro 195 and flanked by extended structure. The last helix, αC, is the longest and is separated from helix αB by a disordered region of ten residues.

Vps2 helix αC contacts MIT helices α2 and α3 and runs parallel with α3. Leu 64^{Vps4} and Glu 68^{Vps4} form key interactions with Arg 224^{Vps2} in the middle of αC (Fig. 1f, i). The mutation of analogous residues in the MIT domain of human Vps4A significantly decreased CHMP1B binding[15]. This central anchor is flanked by additional important hydrophobic interactions and salt links.

Although the crystal structure of the MIT domain of yeast Vps4 agrees closely with NMR structures of the MIT domains of human Vps4A (ref. 15) and Vps4B (ref. 26), our structure of the complex with Vps2 shows that the interaction with ESCRT-III differs from two models that were proposed on the basis of isolated MIT-domain structures. One model[26] proposed that an ESCRT-III subunit would slot into a shallow indentation between MIT helices α1 and α3, corresponding to the locations of Vps2 helices αA and αB in the Vps2–MIT complex. In contrast, another study proposed that an ESCRT-III helix interacts with MIT helix α3 to complete an overall tetratricopeptide repeat (TPR) fold, with the four MIT–ESCRT-III helices packed into a right-handed solenoid[15]. Our structure shows that Vps2 αC fits between MIT helices α2 and α3, but in an orientation opposite to what it would have in a TPR-like arrangement; that is, helix αC packs parallel with MIT helix α3 (Fig. 1e, g).

Six residues along one face of helix αC interact with the MIT domain (Fig. 1f), burying 562 and 676 $Å^2$ of solvent-accessible surface on Vps4 and Vps2, respectively. Leu 221^{Vps2}, Leu 225^{Vps2}, Leu 228^{Vps2} and the hydrophobic portion of Arg 224^{Vps2} form hydrophobic interactions. Three αC residues participate in salt links with the MIT domain, one in the middle (Arg 224^{Vps2}:Glu 68^{Vps4}) and one on each end of αC (Asp 218^{Vps2}:Arg 57^{Vps4} and Lys 229^{Vps2}:Asp 38^{Vps4}). These six ESCRT-III residues define a (D/E)xxLxxRLxxL(K/R) MIT-interacting

[1]MRC Laboratory of Molecular Biology, Medical Research Council Centre, Cambridge CB2 0QH, UK. [2]Department of Molecular Biology and Genetics and Institute for Cell and Molecular Biology, Biotechnology Building, Cornell University, Ithaca, New York 14853, USA.

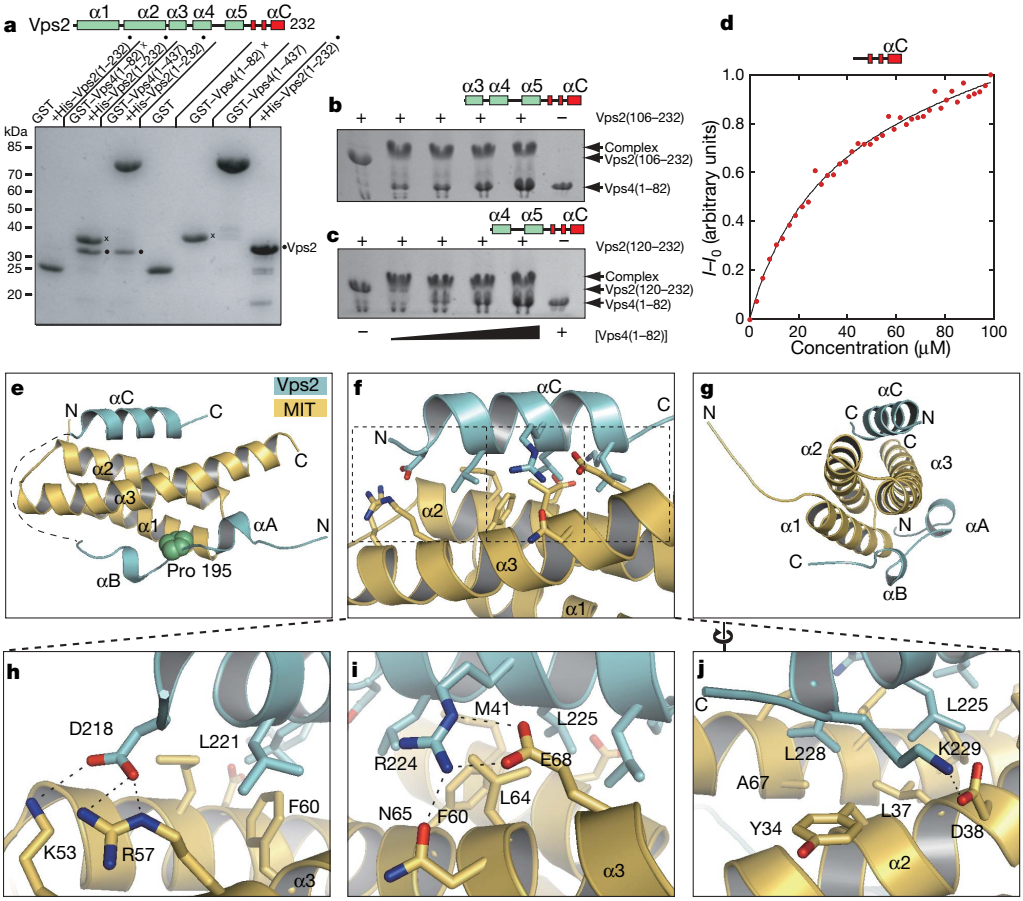


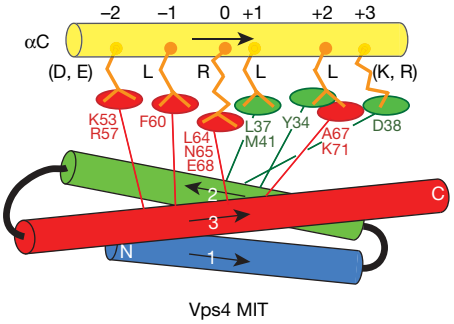
Figure 1 | Characterization of the Vps4–Vps2 complex. **a**, GST-tagged MIT domain of yeast Vps4 and full-length Vps4 interact with full-length Vps2. Coomassie-stained SDS–PAGE shows the material bound to the glutathione-Sepharose resin. **b**, **c**, Vps2 C-terminal constructs (residues 106–232 in **b** and residues 120–232 in **c**) also interact with an untagged MIT domain as detected by band-shift on native PAGE. **d**, Vps2 C-terminal region (residues 183–232) binds to the FlAsH-tagged Vps4 MIT domain with a K_d of 28 μM. **e**, Structure of the complex between Vps2 (cyan) and the Vps4 MIT domain (yellow). **f**, Interactions between Vps2 αC and MIT domain. **g**, The distinctive three-corners-of-a-square appearance of the three-helix MIT bundle. **h**–**j**, Enlarged central (**i**) and peripheral Vps2 helix αC specificity determinants (N-terminal and C-terminal regions are shown in **h** and **j**, respectively).

motif (MIM). We shall refer to these six residues as positions −2, −1, 0, +1, +2 and +3, with the anchoring Arg 224^{Vps2} defined as position 0 (Fig. 2a, b). Point mutations within the Vps2 MIM at positions −1, 0, +1, +2 and +3 eliminate MIT-domain binding *in vitro* (Fig. 2c). MIM residues can be divided into three pairs, a central pair and two peripheral pairs. Within each pair, a hydrophobic residue contributes to affinity and a polar residue contributes to both affinity and specificity of the interaction.

a

```
                  -2 -1  0 +1 +2 +3
                 218    224    229
Sc Vps2    .NPDDDLQARLNTLKKQT      232
Sc Did2    .EKEDKLAQRLRALRG        204
Hs CHMP1A  .SQEDQLSRRLAALRN        196
Hs CHMP1B  .AEQDELSQRLARLRDQV      199
Hs CHMP2A  .DADADLEERLKNLRRD       222
Hs CHMP2B  .ISDEEIERQLKALGVD       213
Hs Vps24   .EALEAMQSRLATLRS        222
Sc Vps24   .RMVNEMRERLRALQN        224
Sc Snf7    .DEEDEDEKALRELQAEMGL    240
Hs CHMP4A  .PKVDEDEEALKQLAEWVS     222
Hs CHMP4B  .KKKEEEDDDMKELENWAGSM   224
Hs CHMP4C  .QRAEEEDDDIKQLAAWAT     233
Hs CHMP5   .DTKNKDGVLVDEFGLPQIPAS  219
Sc Vps60   .EPNLDTEDKNKALESAQ      229
Sc Vps20   .EERSDTKEPLALLLS        221
Hs Vps20   .VKARPRQAELVAAS         201
```

b



c

Mutant	Binding affinity (μM)	*n*
Vps2(183–232)	28.0 ± 4.8	6
Vps2(154–232)	24.5 ± 4.6	3
Vps2(201–232)	29.7 ± 2.1	2
Vps2(208–232)	29.4 ± 1.9	3
Vps2(218–232)	26.8 ± 0.3	2
Vps2(183–232) L190D	32.5 ± 6.9	5
Vps2(183–232) E198R	20.0 ± 1.8	3
Vps2(183–232) T199D	28.1 ± 2.6	3
Vps2(183–232) L221D	No binding	
Vps2(183–232) R224D	No binding	
Vps2(183–232) L225D	No binding	
Vps2(183–232) L228D	No binding	
Vps2(183–232) K229D	No binding	

d

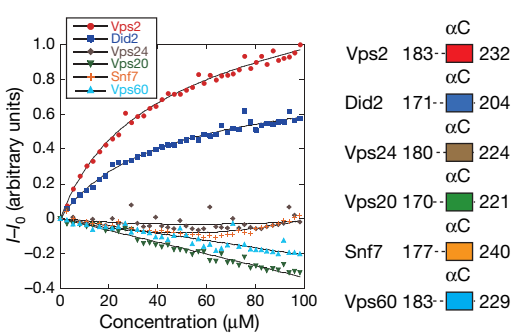


Figure 2 | A conserved MIT-interacting motif in ESCRT-III is critical for binding the Vps4 MIT domain. **a**, Alignment of the C-terminal regions of yeast (Sc) and human (Hs) ESCRT-III subunits. Vps2, Did2 and Vps24 homologues form a distinct group. Residues analogous to Vps2 αC residues contacting Vps4 MIT domain are highlighted. Snf7, Vps20 and Vps60 homologues are more divergent. **b**, Diagram of the ESCRT-III (D/E)xxLxxRLxxL(K/R) motif (yellow). **c**, Binding affinities of Vps2 mutants for MIT domain. Mutations in the MIM impair binding, and helix αC is sufficient for binding. Errors indicate s.d. **d**, Affinities of C-terminal regions from the six ESCRT-III subunits were determined by titration into FlAsH-tagged Vps4 MIT domains. Vps2 and Did2 bind with K_d values of 28 and 39 μM, respectively, and only they have MIMs. No other ESCRT-III binds at concentrations up to 100 μM.

Arg 224^{Vps2} uses both polar guanidinium and aliphatic portions of its side chain to anchor into a central, predominantly hydrophobic socket on the MIT domain made up of Met 41, Phe 60, Leu 64, Asn 65 and Glu 68 (Fig. 1i), forming a salt link with Glu 68^{Vps4}, a hydrogen bond with the Asn 65^{Vps4} side chain, and hydrophobic contacts with Leu 64^{Vps4}. The adjacent Leu 225^{Vps2} fills the bulk of the central, hydrophobic MIT-domain socket. The socket character is conserved between Vps4 MIT domains; however, one of the hydrogen-bonded partners for the central Arg varies (Asn 65 for yeast Vps4 and Asp for Vps4A and Vps4B). This conservation is consistent with the fact that mammalian orthologues are able to complement yeast Vps4 (ref. 27).

Amino-terminal residues of Vps2 αC (MIM positions −1 and −2) interact with MIT α3 (Fig. 1h). The hydrophobic residue at position −1 (Leu 221^{Vps2}) packs against Vps4 Phe 60. The acidic residue at position −2 (Asp 218^{Vps2}) forms salt links with MIT Lys 53^{Vps4} and Arg 57^{Vps4}. At the C-terminal end of helix αC, another pair of residues forms the third set of ESCRT-III–MIT interactions (Fig. 1j), involving primarily MIT helix α2. The hydrophobic residue at the +2 position (Leu 228^{Vps2}) packs against Vps4 Tyr 30, Tyr 34 and Leu 37 in helix α2 and against Ala 67 in helix α3. The +3 basic residue (Lys 229^{Vps2}) interacts with an acidic residue from Vps4 MIT helix α2 (Asp 38^{Vps4}).

Among yeast ESCRT-III-like subunits, only Vps2 and Did2 have the MIM motif (Fig. 2a). Consistent with this is our observation that the MIT domain of Vps4 binds only to Vps2 (K_d 28 μM) and Did2 (K_d 39 μM) (Fig. 2d). Searching the sequence database for MIM motifs identifies Vps2, Did2, CHMP1A and CHMP2A. CHMP1B has a single deviation from this motif, a Gln at position −2 (Gln 185), instead of an acidic residue, but it has a Glu immediately preceding it that may form a similar interaction. Indeed, CHMP1B binds human Vps4A MIT domain with a K_d of 20 μM (ref. 15). CHMP2B, which interacts very poorly with Vps4 (refs 10, 11), lacks the central Arg and has no peripheral basic residues. Yeast Vps24 lacks both the acidic residue at −2 and the basic residue at +3, and does not bind Vps4 (Fig. 2a). Human Vps24 (CHMP3) has a hydrophobic residue instead of an acidic residue at −2, and Met instead of Leu at position −1. These differences might weaken its interaction with Vps4 and explain why there have been conflicting reports about the interaction between Vps4 and human Vps24 (refs 10, 11, 13, 21).

To test whether MIMs have a role in sorting, we mutated motifs in both Vps2 and Did2. Single mutations of either Vps2 (R224D) or Did2 (R198D) or a double Did2 mutant, L199D/L202D, cause partial sorting defects, resulting in mislocalization of green fluorescent protein (GFP)–carboxypeptidase S (CPS) to the limiting membrane of the vacuole. In contrast, a double Vps2 mutant, L228D/K229D, causes a severe sorting defect manifested by a class E compartment (Fig. 3). These results show that MIM motifs are crucial for sorting in yeast.

Vps2 helices αA and αB contact the MIT domain, burying 659 and 723 Å^2 of solvent-accessible surface on Vps4 and Vps2, respectively. Pro 195, separating helices αA and αB, inserts into a conserved MIT pocket. One wall of this pocket is formed by a conserved intramolecular salt link between Asp 21 and Arg 66. However, we find that αA and αB have no influence on MIT binding (Fig. 2c), suggesting that their interactions are due to crystal packing. Consistent with this, there is no sequence similarity between Vps2 and Did2 in the region corresponding to αA and αB.

Other AAA ATPases such as spastin and spartin also have MIT domains, which might be important for their cellular localization[28]. To understand the role of MIT domains in restricting ATPase targets, we compared structural features in more distantly related ATPases. Towards this end, we determined the structure (Supplementary Table 2) of the MIT domain of an archaeal ATPase from *Sulfolobus solfataricus* (SSO0909), which has been annotated as a 'p60 katanin-like' ATPase. The MIT fold is closely conserved, with a Cα root mean squared deviation of only 1.8 Å between the yeast and archaeal MIT domains (Fig. 4). Most conserved features involve hydrophobic packing in the interior, but there is one key salt link, analogous to the Asp 21/Arg 66 link in yeast Vps4, that is universally conserved (Fig. 4a–d). All MIT domains appear as a four-helix bundle lacking one helix. In the archaeal and spartin MIT domains, the slot between helices α2 and α3 lacks specificity determinants important for MIM binding (Fig. 4c). The area analogous to the Vps4 socket that interacts with the central anchor region of the MIM has a basic residue in place of the Leu64^{Vps4} affinity determinant (Lys 65 in *S. solfataricus* ATPase and Arg 89 in spartin). This may reflect a binding surface that evolved to recognize a different target MIM. Structural similarity of archaeal and Vps4 MIT domains, together with the observation that Archaea have ESCRT-III-like subunits (see below), suggests that the archaeal ATPase is more Vps4-like than katanin-like.

All clades of eukaryotes have ESCRTs[29]. All Archaea that have a Vps4-like ATPase also have ESCRT-III-like homologues that form a closely related group (Supplementary Fig. 2), although they do not seem to have ESCRT-0, ESCRT-I or ESCRT-II. Moreover, one ESCRT-III-like subunit is always immediately adjacent to the gene for the Vps4-like ATPase (in *S. solfataricus*, the adjacent ESCRT-III-like subunit is SSO0910). The conserved tandem arrangement of these genes suggests that they might encode functional partners. Indeed, we find that a C-terminal region of SSO0910 binds directly to the archaeal Vps4-like ATPase MIT domain, and residues 164–207 of the ESCRT-III-like protein are both necessary and sufficient for

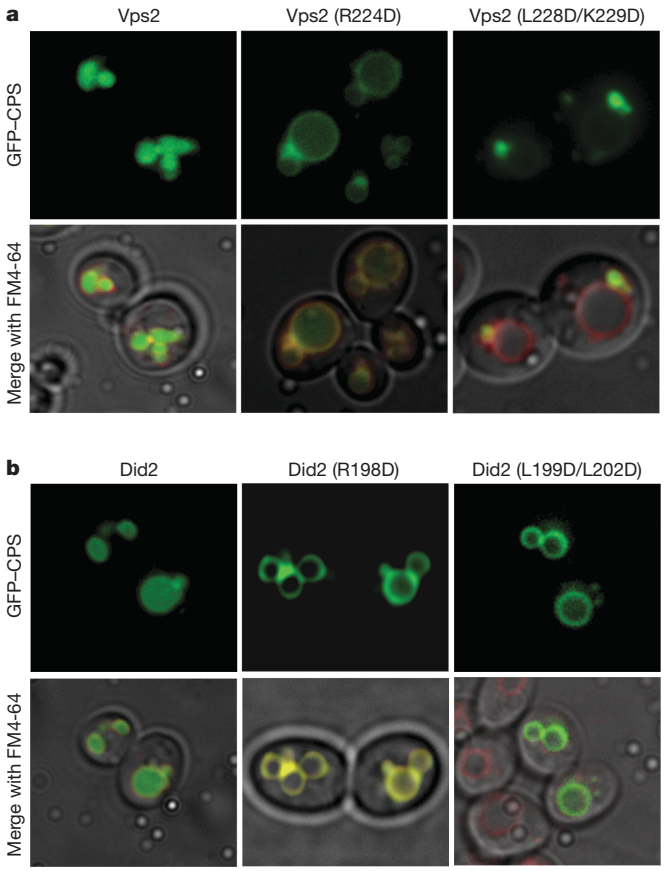


Figure 3 | MIT-interacting motifs in Vps2 and Did2 are important for function *in vivo*. **a**, In wild-type cells, GFP-tagged carboxypeptidase-S accumulates in the vacuolar lumen (FM4-64 preferentially labels the limiting membrane of the vacuole). A single mutation at position 0 of the Vps2 MIM (R224D) impairs sorting, and the GFP–CPS accumulates in the limiting membrane of the vacuole. A double mutation of the Vps2 MIM (L228D/K229D) causes GFP–CPS accumulation in a class E compartment. **b**, Both a single mutation at position 0 of the Did2 MIM (R198D) and a double mutation (L199D/L202D) impair sorting of GFP–CPS, which accumulates in the limiting membrane of the vacuole.

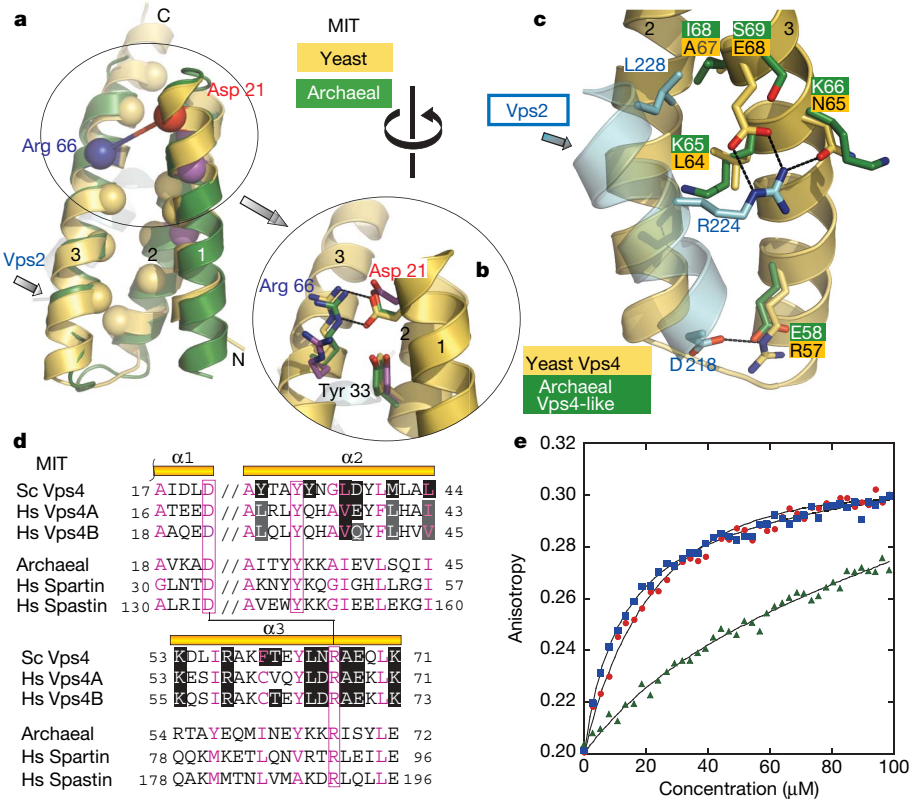


Figure 4 | Conserved features of the MIT domain in ATPases. **a**, Superposition of yeast Vps4 (yellow) and *S. solfataricus* (green) MIT domains. Yellow spheres mark conserved core hydrophobic residues and magenta spheres mark small (Gly or Ala) residues. **b**, Close-up of conserved salt link (including spartin MIT domain (magenta); PDB 2DL1). **c**, Residues in the Vps4–Vps2 interface that differ from the archaeal Vps4-like ATPase. **d**, Selected MIT domains, with conserved residues shown in magenta and ESCRT-III interface residues blackened. Sc, yeast; Hs, human. **e**, Binding of archaeal ESCRT-III-like (SSO0910) C-terminal constructs (titrants) to the archaeal Vps4-like MIT domain, determined by fluorescence anisotropy. Residues 164–207 (blue squares) are both necessary and sufficient (K_d 15 μM). Residues 164–259 (red circles) bind with similar affinity (K_d 20 μM), whereas residues 211–259 (green triangles) do not interact ($K_d > 100$ μM).

this (Fig. 4e). As expected, given the divergence of the archaeal MIT surface (Fig. 4c, d), archaeal ESCRT-III does not bind yeast Vps4 MIT domain (data not shown). The function of the ATPase and ESCRT-III-like components in *S. solfataricus* and other Crenarchaeota is unknown, but Archaea have no endomembrane system, suggesting that ESCRT-III and Vps4 may be relics of a more ancient function common to both Eukaryota and Archaea.

As observed for target binding by other AAA ATPases[30], the isolated Vps4 MIT domain has only a modest affinity for its target ESCRT-III. However, a dodecameric Vps4 should bind an ESCRT-III lattice with a great avidity. Both Vps4-interacting ESCRT-III-related subunits, Vps2 and Did2, are required for efficient vacuolar sorting[6]. Furthermore, the N terminus of Did2 interacts directly with the Vps2–Vps24 subcomplex. Association of the two MIT domain interactors in the same ESCRT-III lattice could enhance the avidity for Vps4. On the ATP hydrolysis-induced transition of Vps4 to a monomer, the low affinity of a single MIT domain for ESCRT-III would facilitate the rapid dissociation of Vps4 from ESCRTs.

The ESCRT-III lattice triggers its own disassembly. It recruits monomeric or dimeric Vps4 to membranes on which the multimeric ESCRT-III lattice facilitates the assembly of the active Vps4 oligomer. Once the ATPase rings have assembled, ATP hydrolysis would lead to the disassembly of both the ATPase rings and the ESCRT-III lattice (Supplementary Fig. 1).

METHODS SUMMARY

The ESCRT-III constructs and the yeast MIT domain were cloned and expressed in *Escherichia coli* with an N-terminal glutathione *S*-transferase (GST) tag. The proteins were purified on glutathione-Sepharose and cleaved with TEV protease, which leaves a four-residue extension, GSHM, at the N terminus. The archaeal MIT domain was expressed with an N-terminal His_6 tag and purified by metal-affinity chromatography. All proteins were further purified by gel filtration. For the crystal structure determinations, Se-Met-substituted proteins were expressed in B834(DE3) cells. For binding measurements, the MIT domains were expressed with a fluorescein bis-arsenical helix binder tag (FlAsH) and fluorescently labelled with Lumio Green (Invitrogen). Binding affinities were determined by titrating protein into a cuvette containing FlAsH-tagged MIT domain and by measuring either a change in fluorescence intensity (yeast MIT domain) or anisotropy (archaeal MIT domain). To enhance the anisotropy change produced by the binding of peptide to the archaeal MIT domain, peptides were titrated as GST fusions.

For the complex of the Vps4 MIT domain with Vps2, the two components were purified separately and mixed for crystallization. The crystal structures were determined by using multiple-wavelength anomalous diffraction (MAD) methods with synchrotron data collected from crystals at 100 K.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 May; accepted 15 August 2007.

1. Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure—diverse function. *Genes Cells* **6**, 575–597 (2001).
2. Slagsvold, T., Pattni, K., Malerod, L. & Stenmark, H. Endosomal and non-endosomal functions of ESCRT proteins. *Trends Cell Biol.* **16**, 317–326 (2006).
3. Williams, R. L. & Urbe, S. The emerging shape of the ESCRT machinery. *Nature Rev. Mol. Cell Biol.* **8**, 355–368 (2007).
4. Bieniasz, P. D. Late budding domains and host proteins in enveloped virus release. *Virology* **344**, 55–63 (2006).
5. Carlton, J. G. & Martin-Serrano, J. Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. *Science* **316**, 1908–1912 (2007).
6. Nickerson, D. P., West, M. & Odorizzi, G. Did2 coordinates Vps4-mediated dissociation of ESCRT-III from endosomes. *J. Cell Biol.* **175**, 715–720 (2006).

7. Babst, M., Wendland, B., Estepa, E. J. & Emr, S. D. The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J.* **17,** 2982–2993 (1998).
8. Howard, T. L., Stauffer, D. R., Degnin, C. R. & Hollenberg, S. M. CHMP1 functions as a member of a newly defined family of vesicle trafficking proteins. *J. Cell Sci.* **114,** 2395–2404 (2001).
9. Amerik, A., Nowak, J., Swaminathan, S. & Hochstrasser, M. The Doa4 deubiquitinating enzyme is functionally linked to the vacuolar protein-sorting and endocytic pathways. *Mol. Biol. Cell* **11,** 3365–3380 (2000).
10. Tsang, H. T. *et al.* A systematic analysis of human CHMP protein interactions: Additional MIT domain-containing proteins bind to multiple components of the human ESCRT III complex. *Genomics* **88,** 333–346 (2006).
11. von Schwedler, U. K. *et al.* The protein network of HIV budding. *Cell* **114,** 701–713 (2003).
12. Bowers, K. *et al.* Protein–protein interactions of ESCRT complexes in the yeast *Saccharomyces cerevisiae*. *Traffic* **5,** 194–210 (2004).
13. Martin-Serrano, J., Yarovoy, A., Perez-Caballero, D. & Bieniasz, P. D. Divergent retroviral late-budding domains recruit vacuolar protein sorting factors by using alternative adaptor proteins. *Proc. Natl Acad. Sci. USA* **100,** 12414–12419 (2003).
14. Vajjhala, P. R., Catchpoole, E., Nguyen, C. H., Kistler, C. & Munn, A. L. Vps4 regulates a subset of protein interactions at the multivesicular endosome. *FEBS J.* **274,** 1894–1907 (2007).
15. Scott, A. *et al.* Structure and ESCRT-III protein interactions of the MIT domain of human VPS4A. *Proc. Natl Acad. Sci. USA* **102,** 13813–13818 (2005).
16. Bishop, N. & Woodman, P. ATPase-defective mammalian VPS4 localizes to aberrant endosomes and impairs cholesterol trafficking. *Mol. Biol. Cell* **11,** 227–239 (2000).
17. Scott, A. *et al.* Structural and mechanistic studies of VPS4 proteins. *EMBO J.* **24,** 3658–3669 (2005).
18. Babst, M., Katzmann, D. J., Estepa-Sabal, E. J., Meerloo, T. & Emr, S. D. ESCRT-III: an endosome-associated heterooligomeric protein complex required for MVB sorting. *Dev. Cell* **3,** 271–282 (2002).
19. Muziol, T. *et al.* Structural basis for budding by the ESCRT-III factor CHMP3. *Dev. Cell* **10,** 821–830 (2006).
20. Lin, Y., Kimpler, L. A., Naismith, T. V., Lauer, J. M. & Hanson, P. I. Interaction of the mammalian endosomal sorting complex required for transport (ESCRT) III protein hSnf7-1 with itself, membranes, and the AAA+ ATPase SKD1. *J. Biol. Chem.* **280,** 12799–12809 (2005).
21. Zamborlini, A. *et al.* Release of autoinhibition converts ESCRT-III components into potent inhibitors of HIV-1 budding. *Proc. Natl Acad. Sci. USA* **103,** 19140–19145 (2006).
22. Agromayor, M. & Martin-Serrano, J. Interaction of AMSH with ESCRT-III and deubiquitination of endosomal cargo. *J. Biol. Chem.* **281,** 23083–23091 (2006).
23. McCullough, J. *et al.* Activation of the endosome-associated ubiquitin isopeptidase AMSH by STAM, a component of the multivesicular body-sorting machinery. *Curr. Biol.* **16,** 160–165 (2006).
24. Whitley, P. *et al.* Identification of mammalian Vps24p as an effector of phosphatidylinositol 3,5-bisphosphate-dependent endosome compartmentalization. *J. Biol. Chem.* **278,** 38786–38795 (2003).
25. Yorikawa, C. *et al.* Human CHMP6, a myristoylated ESCRT-III protein, interacts directly with an ESCRT-II component EAP20 and regulates endosomal cargo sorting. *Biochem. J.* **387,** 17–26 (2005).
26. Takasu, H. *et al.* Structural characterization of the MIT domain from human Vps4b. *Biochem. Biophys. Res. Commun.* **334,** 460–465 (2005).
27. Scheuring, S. *et al.* Mammalian cells express two VPS4 proteins both of which are involved in intracellular protein trafficking. *J. Mol. Biol.* **312,** 469–480 (2001).
28. Reid, E. *et al.* The hereditary spastic paraplegia protein spastin interacts with the ESCRT-III complex-associated endosomal protein CHMP1B. *Hum. Mol. Genet.* **14,** 19–38 (2005).
29. Field, M. C., Gabernet-Castello, C. & Dacks, J. B. in *Origins and Evolution of Eukaryotic Endomembranes and Cytoskeleton* (ed. Jékely, G.) 84–96 (Eurekah/Landes Bioscience Press, Austin, TX, 2007).
30. Hartman, J. J. & Vale, R. D. Microtubule disassembly by ATP-dependent oligomerization of the AAA enzyme katanin. *Science* **286,** 782–785 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank M. Babu for advice on the archaeal genome analysis. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities and we thank G. Cioci, D. Flot, I. Leiros and G. Leonard for assistance in using beamlines ID23-2 and ID23-1. T.O. was supported by a JSPS fellowship and S.S. by a fellowship from the Howard Hughes Medical Institute. This research was supported by the Howard Hughes Medical Institute (S.D.E.) and the Medical Research Council (R.L.W.).

Author Information The atomic coordinates of the yeast Vps2–Vps4 MIT-domain complex and the *S. solfataricus* MIT domain are deposited in the Protein Data Bank under accession numbers 2V6X and 2V6Y, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.L.W. (rlw@mrc-lmb.cam.ac.uk).

METHODS

Protein cloning, expression and purification. Yeast Vps4 MIT domain (residues 1–82), full-length yeast Vps4 (residues 1–437), Vps2 C terminus (residues 183–232), Snf7 C terminus (residues 177–240), Vps20 C terminus (residues 170–221), Vps24 C terminus (residues 180–224), Did2 C terminus (residues 171–204) and Vps60 C terminus (residues 183–229) were cloned with an N-terminal GST tag in the pOPTG vector and expressed in C41(DE3)RIPL cells. Full-length Vps2 used for pull-down experiments was expressed with an N-terminal MAHHHHHH tag. For crystallography, the Se-Met-substituted protein was expressed in a B834(DE3) methionine auxotroph. Cells were lysed in buffer A (20 mM Tris-HCl, pH 8.0 at 4 °C, 100 mM NaCl and 1 mM dithiothreitol (DTT)) and incubated with glutathione-Sepharose 4B (GE Healthcare) for 1 h. After being washed, the GST fusion protein was digested with TEV protease (100:1 w/w) at 4 °C for 12 h on the resin. The TEV-eluted protein was further purified by gel filtration on a Superdex 75 16/60 column in buffer B (Tris-HCl, pH 7.4 at 20 °C, 100 mM NaCl and 1 mM DTT). Vps2 mutants were also cloned with an N-terminal GST tag in the pOPTG vector, expressed in C41(DE3) cells and purified with the same methods as for the Vps2 C terminus (residues 183–232). *Sulfolobus solfataricus* ESCRT-III-like protein (SSO0910) constructs (residues 164–259, 164–207 and 211–259) were amplified by polymerase chain reaction (PCR) from genomic DNA, expressed and purified similarly to Vps2 constructs except that the GST tag was not cleaved. All constructs were verified by sequencing.

The His_6FlAsH-tagged yeast Vps4 MIT domain was expressed by using the pOPTHF vector in C41(DE3)RIPL cells. Cells were lysed in buffer C (20 mM Tris-HCl pH 8.0, 20 mM imidazole, 100 mM NaCl and 2 mM 2-mercaptoethanol). The protein was purified on a 5-ml His-Trap-FF column equilibrated with the same buffer. The column was washed in the same buffer with 0.1% Triton X-100 and the protein was eluted with an imidazole gradient. Subsequent purification was conducted on a HiTrap-Q column (GE Healthcare) in buffer D (20 mM Tris-HCl, pH 8.5 at 20 °C, 1 mM DTT) and a NaCl gradient. The eluted protein was digested with TEV protease (100:1 w/w) at 4 °C for 12 h to remove the His_6 tag, then purified by gel filtration on a Superdex 75 16/60 column in buffer E (Tris-HCl, pH 7.4 at 20 °C, 100 mM NaCl and 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP)). FlAsH-tagged Vps4 MIT domain (10 nmol) was labelled with 10 nmol of Lumio Green detection reagent (Invitrogen) in buffer E (containing 5 mM 2-mercaptoethanol) in a final reaction volume of 1 ml for 2 h at 4 °C. The sample was dialysed overnight against buffer E (containing 5 mM 2-mercaptoethanol) at 4 °C with a 3.5-kDa Slide-A-Lyzer membrane (Pierce). The His_6FlAsH-tagged full-length Vps2 (residues 1–232) and the archaeal MIT domain were also expressed by using the pOPTHF vector and purified similarly to the His_6FlAsH-tagged Vps4 MIT domain except that the His_6-tag was not cleaved with TEV.

The MIT domain from the N terminus of the *S. solfataricus* Vps4-like ATPase (accession number SSO0909) was PCR amplified from genomic DNA and cloned into a pOPTH vector, encoding an N-terminal MAHHHHHH tag. The Se-Met-substituted protein was expressed in B834(DE3) cells at 37 °C for 3 h after induction with isopropyl β-D-thiogalactoside at a D_{600} of 1.0. The protein was purified by Ni-affinity chromatography, anion exchange and gel filtration (20 mM Tris-HCl pH 8, 100 mM NaCl, 2 mM DTT).

Yeast plasmid construction and yeast strains. Yeast plasmids and strains are described in Supplementary Methods.

Microscopy. Living cells expressing the GFP–CPS chimaera were harvested at a D_{600} of 0.4–0.7, labelled with FM4-64 for vacuolar membrane staining and resuspended in medium for visualization. Visualization of cells was performed on a fluorescence microscope (Axiovert S1002TV; Carl Zeiss MicroImaging) equipped with fluorescein isothiocyanate and rhodamine filters, captured with a digital camera (CH350 CCD; Photometrix), and deconvolved with Delta Vision software (Applied Precision). Results presented were based on observations of more than 120 cells.

Crystallization. The LMB nanolitre crystallization robotic facility was used for a broad initial screen of 1,440 crystallization conditions. Optimal crystals for the complex of the yeast Vps4 MIT domain (residues 1–82) with Vps2 C terminus (residues 183–232) (both cleaved from GST fusions) were obtained at 17 °C by vapour diffusion from a protein solution at 28 mg ml^{-1} containing a 1:1.7 molar ratio of MIT:Vps2 and a reservoir solution containing 2.0 M $(NH_4)_2SO_4$ and 0.1 M HEPES (pH 7.0). Crystals were cryoprotected by adding glycerol to a final concentration of 25% and frozen by immersion in liquid nitrogen.

The *S. solfataricus* MIT domain was crystallized in sitting drops by vapour diffusion by mixing 2 μl of reservoir solution (0.6 M ammonium tartrate and 2% PEG4K) with 1 μl of protein solution (9 mg ml^{-1} in gel filtration buffer). Crystals appeared within three days after incubation at 17 °C. For data collection, a crystal in a MicroMount loop (MiTeGen) was placed on a Proteros free-mounting system, and all excess liquid was then removed from the loop with a piece of filter paper. The loop was then rapidly immersed in liquid nitrogen.

Crystallographic structure determination. Data for the Vps2–Vps4 complex were collected at 100 K on ESRF beamlines ID23-1 and ID23-2 with the use of two crystals (see Supplementary Methods). The figure-of-merit for the Se-Met MAD refinement after AutoSHARP was 0.3 and after Density Modification (DM) was 0.92. The Ramachandran plot had 96.4% of residues in the core regions and none disallowed. Data collection and refinement statistics are given in Supplementary Tables 1 and 2.

Se-MET MAD data for the archaeal MIT domain were collected at 100 K on ESRF beamline ID23-1. The figure-of-merit after AutoSHARP was 0.44 and after DM was 0.85. The Ramachandran plot had 94.9% of residues in the core regions and none disallowed.

Fluorescence titration binding assay. Analyte protein was titrated into a cuvette containing 15 nM Vps4 MIT domain, 33 nM full-length Vps2 or 20 nM archaeal MIT domain N-terminally labelled with Lumio Green (FlAsH-MIT and FlAsH-Vps2) in 1.1 ml of binding buffer (20 mM Tris-HCl, pH 7.4 at 20 °C, 100 mM NaCl and 5 mM 2-mercaptoethanol). Fluorescence was measured with a Perkin-Elmer LS-55 spectrophotometer with an excitation wavelength of 490 nm and an emission wavelength of 530 nm. Excitation and emission slits were either 10 nm or 15 nm and 20 nm, respectively. A 1.0–1.5-mM protein analyte was titrated into a cuvette with a MicroLab 500 titrator (Hamilton). The K_d values were calculated from a direct fitting of the titration data to a single-site model. At least two independent experiments were conducted to determine K_d values.

LETTERS

ESCRT-III recognition by VPS4 ATPases

Melissa D. Stuchell-Brereton[1]*, Jack J. Skalicky[1]*, Collin Kieffer[1], Mary Anne Karren[1], Sanaz Ghaffarian[1] & Wesley I. Sundquist[1]

The ESCRT (endosomal sorting complex required for transport) pathway is required for terminal membrane fission events in several important biological processes, including endosomal intraluminal vesicle formation[1,2], HIV budding[3] and cytokinesis[4]. VPS4 ATPases perform a key function in this pathway by recognizing membrane-associated ESCRT-III assemblies and catalysing their disassembly[5–7], possibly in conjunction with membrane fission. Here we show that the microtubule interacting and transport (MIT) domains of human VPS4A and VPS4B bind conserved sequence motifs located at the carboxy termini of the CHMP1–3 class of ESCRT-III proteins. Structures of VPS4A MIT–CHMP1A and VPS4B MIT–CHMP2B complexes reveal that the C-terminal CHMP motif forms an amphipathic helix that binds in a groove between the last two helices of the tetratricopeptide-like repeat (TPR) of the VPS4 MIT domain, but in the opposite orientation to that of a canonical TPR interaction. Distinct pockets in the MIT domain bind three conserved leucine residues of the CHMP motif, and mutations that inhibit these interactions block VPS4 recruitment, impair endosomal protein sorting and relieve dominant-negative VPS4 inhibition of HIV budding. Thus, our studies reveal how the VPS4 ATPases recognize their CHMP substrates to facilitate the membrane fission events required for the release of viruses, endosomal vesicles and daughter cells.

MIT domains are found in a number of proteins involved in cellular trafficking, where they frequently function as protein interaction modules[8,9]. The ESCRT-III proteins, in particular, interact with a series of different MIT-domain-containing proteins[10] including VPS4A and VPS4B[6,11], spastin[12], AMSH (associated molecule with the SH3 domain of STAM)[10,13–15], and others[10]. Biosensor binding experiments demonstrated that the human VPS4B MIT domain bound both full-length CHMP1B and a shorter, more soluble construct ($CHMP1B_{65–196}$) with dissociation constants of about 30 μM (Fig. 1a, dark blue and red curves)[11]. A series of amino-terminal deletion constructs were used to map the minimal VPS4B MIT-binding site on CHMP1B, and these experiments revealed that a construct spanning just the C-terminal 17 residues of CHMP1B bound the VPS4B MIT domain with nearly full affinity (Fig. 1a, green curve; K_d 34 ± 1 μM (mean ± s.d.). In contrast, a CHMP1B construct lacking this region ($CHMP1B_{65–180}$) did not bind (Fig. 1a, black curve), in good agreement with reports that deletion of the C terminus of CHMP3 blocks VPS4 and AMSH binding[14,16]. Similar binding data were obtained for the VPS4A MIT domain (not shown), indicating that the C-terminal end of CHMP1B is necessary and sufficient for VPS4 MIT binding.

The ten canonical human ESCRT-III proteins fall into two general classes (CHMP1–3 and CHMP4–6) based on sequence homology and functional similarities. These two classes can be further subdivided into the six different CHMP families, each of which corresponds to one of the six ESCRT-III-like proteins in yeast[7,17,18]. Sequence alignments revealed that three highly conserved leucine residues are arrayed in a heptad repeat at the C termini of CHMP1–3 proteins but not CHMP4–6 proteins (Fig. 2a, CHMP1A positions 187, 191 and 194). Mutation of either of the two most conserved leucine residues within CHMP1B decreased VPS4A and VPS4B MIT binding more than tenfold, confirming the importance of these residues for interactions between VPS4 MIT and CHMP1B (Fig. 1a, orange and teal curves and data not shown).

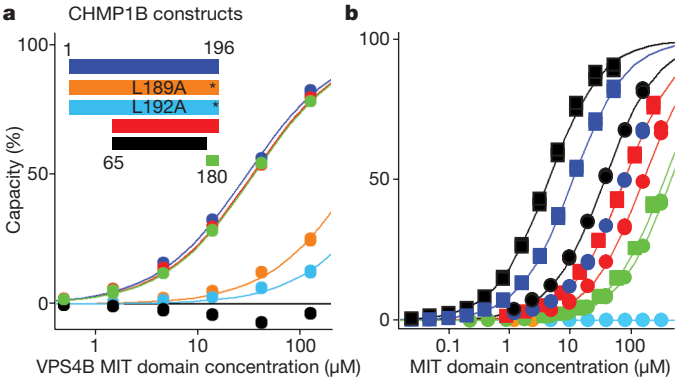


Figure 1 | VPS4 MIT–CHMP binding interactions. a, Biosensor binding isotherms showing VPS4B MIT binding to the wild-type and mutant CHMP1B proteins depicted schematically in the inset. The data demonstrate that the final 17 residues of CHMP1B are necessary and sufficient for full-affinity VPS4B MIT binding and that binding is inhibited by mutations in the two most conserved residues within this region (L189A and L192A). Duplicate data points are shown for each condition, and estimated dissociation constants derived from fits to single site binding models (curves) were as follows: CHMP1B, 29 ± 1 μM; $CHMP1B_{65–196}$, 33 ± 1 μM; $CHMP1B_{180–196}$, 34 ± 1 μM; $CHMP1B_{L189A}$, 353 ± 8 μM; $CHMP1B_{L192A}$, 810 ± 40 μM (means ± s.d.). Note that in other cases, VPS4 MIT domains bound even more tightly to isolated binding sites than to the corresponding full-length CHMP proteins, which is consistent with an autoinhibitory mechanism (refs 16, 29, 30; data not shown). **b**, Biosensor binding isotherms showing C-terminal peptides from CHMP1A and CHMP2B binding GST–VPS4A MIT and GST–VPS4B MIT under different solution conditions. The data show that both VPS4 MIT domains bind specifically to the CHMP1A and CHMP2B C termini, and that the affinities vary with different MIT domains, CHMP proteins and solution conditions. Representative data are shown for low salt and pH conditions ('L', 20 mM phosphate pH 5.5 (NMR conditions); squares) and medium salt and pH conditions ('M', 20 mM Tris-HCl pH 8.0, 100 mM NaCl; circles). Estimated dissociation constants derived from fits to single-site binding models were as follows: VPS4A MIT–$CHMP1A_{180–196}$ (black), 4.61 ± 0.02 μM (L) and 33.4 ± 0.2 μM (M); VPS4A MIT–$CHMP2B_{195–213}$ (red), 71.8 ± 0.3 μM (L) and 178 ± 1 μM (M); VPS4B MIT–$CHMP1A_{180–196}$ (blue), 13.4 ± 0.1 μM (L) and 74.0 ± 0.3 μM (M); VPS4B MIT–$CHMP2B_{195–213}$ (green), 273 ± 0.2 μM (L) and 402 ± 2 μM (M). Background binding of both CHMP peptides to control GST surfaces was negligible under both conditions, as shown for the medium-salt conditions ($CHMP1A_{180–196}$, orange; $CHMP2B_{195–213}$, cyan).

[1]Department of Biochemistry, Room 4100, 15 N. Medical Drive East, University of Utah, Salt Lake City, Utah 84112-5650, USA.
*These authors contributed equally to this work.

To assess the generality of VPS4 MIT binding to the C termini of the CHMP1–3 proteins, peptides corresponding to these regions of CHMP1A, CHMP1B, CHMP2A and CHMP2B were tested and shown to bind the MIT domains of VPS4A and VPS4B in every case (Fig. 1b, Supplementary Fig. 1, and data not shown). Thus, this sequence motif constitutes a generalized VPS4 MIT-binding site. In contrast, binding of VPS4 MIT to proteins of the CHMP4–6 class was much more variable and the binding site did not map to the C terminus of CHMP6 (the one case studied in detail). VPS4 MIT domains can therefore also bind a subset of the CHMP4–6 proteins, but the recognition mode is different and will be described elsewhere.

Our initial surveys revealed that dissociation constants for MIT binding to CHMP1A, CHMP1B, CHMP2A and CHMP2B peptides varied throughout the micromolar range, with tighter binding generally associated with VPS4A, with reduced temperatures, with reduced ionic strengths and with slightly acidic conditions. Several of these trends are evident in the isotherms for VPS4 MIT domains binding to $CHMP1A_{180–196}$ and $CHMP2B_{195–213}$ (Fig. 1b). Representative strong (VPS4A MIT–$CHMP1A_{180–196}$; K_d 4.6 μM) and weak (VPS4B MIT–$CHMP2B_{195–213}$; K_d 273 μM) complexes were selected for structural studies by NMR. Studies of VPS4A MIT–$CHMP1A_{180–196}$ and free VPS4B MIT were straightforward and yielded high-quality structures (Fig. 2, Supplementary Figs 1–3 and Supplementary Table 1), whereas assignment of intermolecular nuclear Overhauser effects in the weaker VPS4B MIT–$CHMP2B_{195–213}$ complex required reference to the VPS4A MIT–$CHMP1A_{180–196}$ structure (Supplementary Fig. 4, Supplementary Table 1 and Supplementary Methods). Thus, although the two VPS4 MIT–CHMP structures were similar, only the VPS4A MIT–$CHMP1A_{180–196}$ complex is described here in detail.

N-terminal domains of AAA ATPases generally dictate substrate specificity and therefore vary considerably in structure[19]. The VPS4A MIT domain is a three-helix bundle[11,20], and $CHMP1A_{180–196}$ forms a fourth helix that binds in the groove between helices 2 and 3 (Fig. 2b, c). The three conserved $CHMP1A_{180–196}$ leucine residues line one side of the amphipathic helix, and each binds in a hydrophobic pocket along the groove (Fig. 2c–f). These interactions seem to constitute the primary recognition determinants, but complementary salt bridges are also made by three of the adjacent conserved CHMP1A residues (Glu 184, Arg 190 and Arg 195; see Fig. 2). The structure nicely explains existing mutagenesis data showing the importance of several MIT residues for CHMP binding[11]. In particular, the VPS4A MIT Leu 64 residue has a critical function in helping to define all three MIT-binding pockets (highlighted in blue in Fig. 2c–f), and an L64A mutation decreases CHMP1B binding more than 30-fold[11]. The structures indicated that an Asp residue at this position would weaken binding even further by interacting unfavourably with all three conserved CHMP1–3 leucine residues, and, indeed, VPS4A MIT_{L64D} did not detectably bind glutathione *S*-transferase (GST)-tagged $CHMP1B_{180–196}$ at protein concentrations up to 250 μM (Supplementary Fig. 5). A VPS4A MIT E68A mutation also decreased the CHMP1B binding affinity 2–3-fold[11], which is consistent with the conserved salt bridge observed between VPS4A Glu 68 and CHMP1A Arg 190.

The structures further revealed that two previous proposals for MIT–CHMP interactions were incorrect. Specifically, CHMP1–3 proteins do not bind between MIT helices 1 and 2 to complete a canonical four-helix bundle[20], nor does the binding of the CHMP helix between MIT helices 2 and 3 complete a second paired tetratricopeptide helical repeat[11]. Instead, the CHMP helix binds parallel to MIT helix 3, whereas a TPR-like interaction would have the opposite orientation. Hence, although the TPR-like disposition of the three MIT helices helps to define the geometry of the CHMP binding site, the interaction mode has not previously been observed.

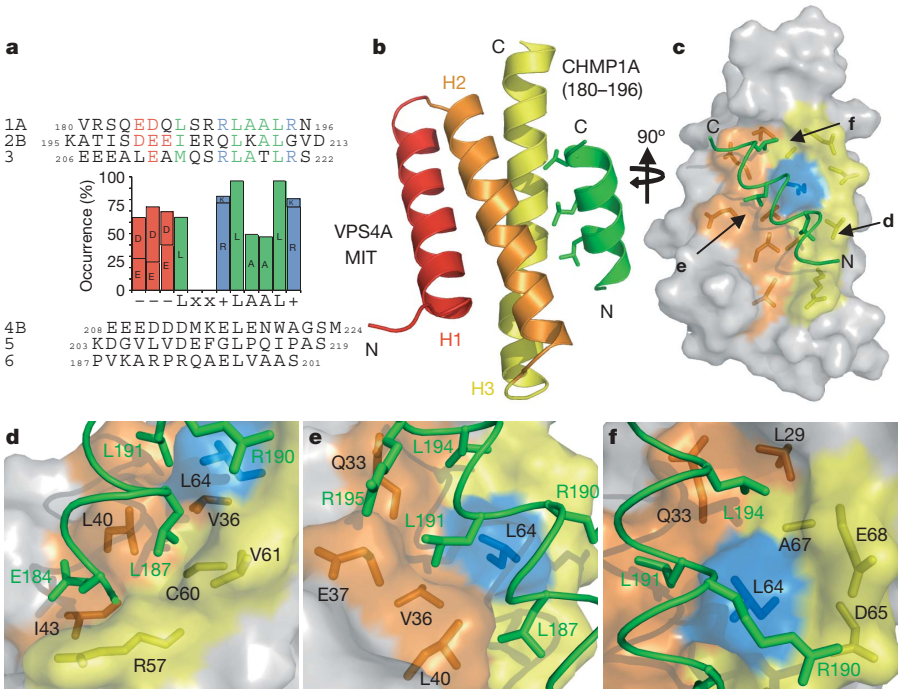


Figure 2 | Structural basis for VPS4 MIT recognition of CHMP1–3. **a**, Alignments of ESCRT-III protein C termini. C-terminal sequences of human proteins from the six CHMP classes are shown explicitly, together with a graph showing the degree of sequence conservation within aligned CHMP1–3 C termini (50 sequences; see Supplementary Table 2) and the consensus sequence (below the graph). Note that ESCRT-III proteins of the CHMP4–6 class do not conform to the consensus. **b**, Solution structure of the VPS4A MIT–$CHMP1A_{180–196}$ complex, with the three conserved $CHMP1A_{180–196}$ leucines shown explicitly. This helical colour scheme is also used in **c**–**f**. **c**, Structure of the VPS4A MIT–$CHMP1A_{180–196}$ complex. The MIT domain is shown in a space-filling model with Leu 64 highlighted in blue; important residues on both sides of the interface are shown explicitly. Arrows denote the approximate orientations of the three leucine-binding pockets shown in **d**–**f**. **d**–**f**, Close-up views of the three leucine-binding pockets of the VPS4A MIT domain showing the hydrophobic pocket views in detail, as well as complementary charge interactions between Glu 184 and Arg 57 (**d**), between Arg 195 and Glu 37 (**e**), and between Arg 190 and Glu 68 (Asp 65) (**f**); CHMP1A residues are listed first.

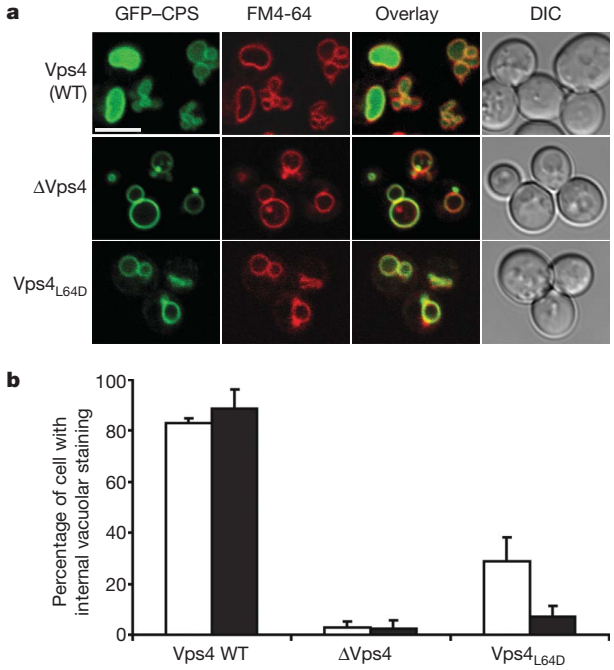


Figure 3 | The Vps4 MIT Leu64Asp mutation inhibits membrane protein sorting into the vacuolar lumen. **a**, Confocal fluorescence slices of live yeast cells showing localization of the model membrane protein cargo, GFP–CPS, in the presence of wild-type (WT) Vps4 (top row), no Vps4 (middle row) or $Vps4_{L64D}$ (bottom row), at 30 °C. FM4-64 staining (red) was used to define the limiting vacuolar membrane (open circles) and to reveal class E compartments (intense puncta in the middle and bottom rows). Overlaid fluorescence and differential interference contrast (DIC) images are shown in the right two columns for reference. Scale bar, 5 μm. **b**, Graphic quantification of the experiment shown in **a**, demonstrating that the $Vps4_{L64D}$ mutant inhibits GFP–CPS trafficking into the lumen of the vacuole. GFP–CPS localization was examined at 30 °C (open columns) and 37 °C (filled columns). Data are from three independent experiments (100 cells per experiment); error bars indicate s.d.

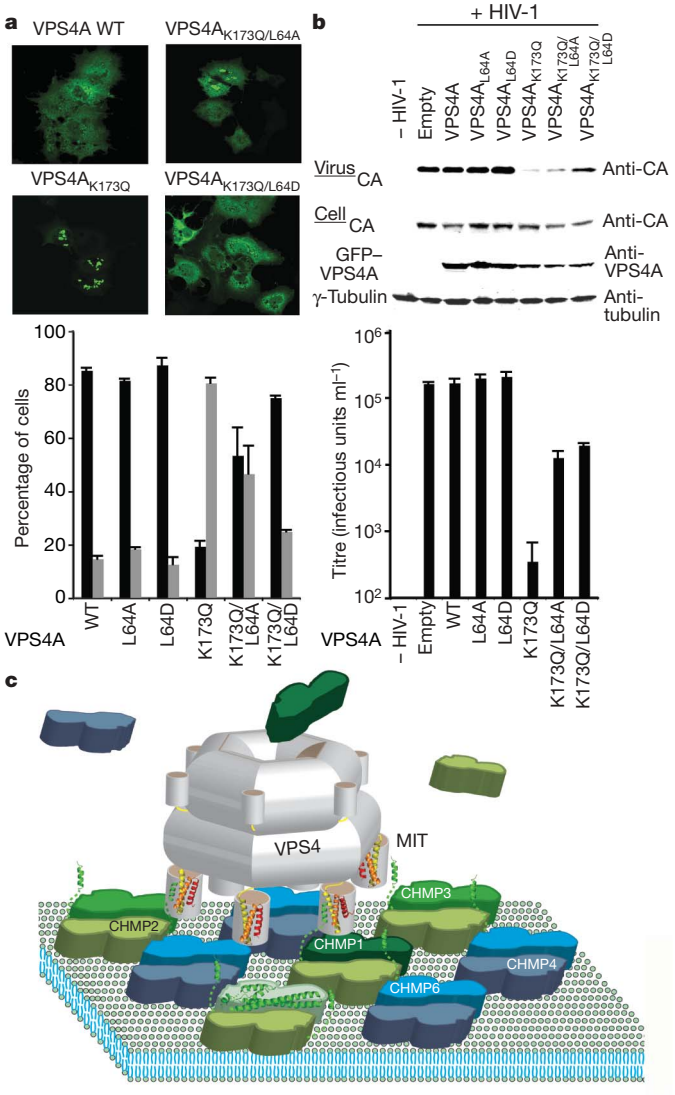


Figure 4 | Secondary MIT Leu 64 mutations decrease endosomal recruitment and inhibition of HIV-1 budding by VPS4A proteins that lack ATP-binding activity. **a**, Upper panels: confocal fluorescence slices showing representative distributions of wild-type (WT) GFP–VPS4A, GFP–$VPS4A_{K173Q}$, GFP–$VPS4A_{K173Q/L64A}$ and GFP–$VPS4A_{K173Q/L64D}$, in transfected COS7 cells. Lower panel: quantification showing that the K173Q mutation causes VPS4A to become trapped on aberrant endosomal membranes (class E compartments) and that secondary mutations at Leu 64 restore the cytoplasmic localization of VPS4A partly (L64A) or almost fully (L64D). Black columns, diffuse staining; grey columns, punctate staining. Data are from two independent experiments (at least 100 cells per experiment); error bars indicate s.d. **b**, Secondary L64A and L64D mutations decrease dominant inhibition of HIV-1 vector release and infectivity by GFP–$VPS4A_{K173Q}$. Lanes 3–8 show HIV-1 vector titres (lower panel) and virion release (upper panels) on co-expression with GFP–VPS4A (lanes 3, negative control), GFP–$VPS4A_{L64A}$ (lanes 4, negative control), GFP–$VPS4A_{L64D}$ (lanes 5), GFP–$VPS4_{K173Q}$ (lanes 6, positive control) or the doubly mutant GFP–$VPS4_{K173Q/L64A}$ (lanes 7) or GFP–$VPS4_{K173Q/L64D}$ (lanes 8) proteins. Virion release (top panel, virus) was assayed by western blot detection of virion-associated CA proteins. The lower western blots show, respectively, cell-associated Gag, exogenous GFP–VPS4A proteins and a γ-tubulin loading control. Controls lacking HIV-1 vector and exogenous GFP–VPS4A are shown in lanes 1 and 2, respectively. Vector titres (infectious units ml^{-1}) are shown in the bottom panel (triplicate measurements; error bars indicate s.d.). **c**, Schematic model showing VPS4 recruitment to the membrane-associated ESCRT-III lattice. An idealized membrane-associated ESCRT-III lattice, based on the CHMP3 crystal structure and lattice contacts[2,29], is shown with the C-terminal tails on the CHMP1–3 subunits exposed to interact with the MIT domains of a dodecameric VPS4 protein (grey). The VPS4 ATPase activity releases the assembled ESCRT-III lattice from the membrane and may also be coupled to vesicle formation (not shown).

Yeast Vps4 is required for the formation of intraluminal vesicles that transport membrane proteins into the late endosome and ultimately into the lumen of the vacuole[5,6]. Loss of Vps4 activity therefore induces the formation of aberrant endosomes (termed class E compartments) and inhibits membrane protein cargoes such as green fluorescent protein-labelled carboxypeptidase S (GFP–CPS) from accumulating within the vacuole (Fig. 3; compare rows 1 and 2). A single L64D point mutation within the Vps4 MIT domain also impaired the vacuolar accumulation of the GFP–CPS cargo significantly at 30 °C and almost entirely at 37 °C, showing that the structurally characterized VPS4 MIT–ESCRT-III interface is required for efficient endosomal protein sorting.

Mutations that block ATP binding by Vps4 (for example $Vps4_{K179Q}$) also induce class E compartment formation; in this case all of the ESCRT machinery, including Vps4 itself, remains trapped on the class E compartments, leading to punctate Vps4 staining in virtually every cell (Supplementary Fig. 6)[5,6]. However, the secondary L64D MIT-domain mutation substantially redistributed the $Vps4_{L64D/K179Q}$ protein from class E compartments back into the cytoplasm (Supplementary Fig. 6). Thus, this mutation inhibits protein sorting at least in part by preventing the recruitment of Vps4 to endosomal membranes.

The importance of the MIT–CHMP interaction for VPS4 recruitment to mammalian endosomes was also tested by comparing the cellular distributions of different exogenously expressed GFP–VPS4A fusion proteins (Fig. 4a). As in yeast, VPS4 ATP-binding mutants (for example $VPS4A_{K173Q}$) induce the formation of mammalian class E compartments that trap VPS4 and other ESCRT

machinery on their extended surfaces (Fig. 4a; compare VPS4A with $VPS4A_{K173Q}$)[6,21,22]. Quantitative analyses revealed that wild-type GFP–VPS4A was distributed diffusely throughout cytoplasm in 85% of cells, whereas GFP–$VPS4A_{K173Q}$ was trapped on punctate class E compartments in 79% of cells (Fig. 4a, lower panel). However, the secondary L64A mutation significantly decreased punctate/class E localization of $VPS4A_{K173Q/L64A}$ (to 47%), and the more detrimental L64D mutation decreased punctate/class E localization even further (25%), to nearly the same level as that of the wild-type protein (15%). The redistribution of doubly mutant VPS4A proteins is particularly significant for two reasons: first because these experiments were performed in the presence of endogenous wild-type VPS4A and VPS4B, which probably co-polymerize with the $VPS4_{K173Q/L64A,D}$ mutants and augment endosomal recruitment, and second because LIP5/Vta1, a VPS4 cofactor, seems to interact with both VPS4 and ESCRT-III proteins and may therefore also contribute to VPS4 recruitment[23–25].

We also tested whether the secondary L64A and L64D mutations decreased the dominant inhibition of HIV-1 release and infectivity by the $VPS4A_{K173Q}$ mutant. As reported previously[26], overexpression of the $VPS4A_{K173Q}$ mutant strongly inhibited HIV-1 release, decreasing infectious titres more than 500-fold (Fig. 4b). However, incorporation of secondary L64A or L64D mutations restored virus release and infectivity 35-fold and 55-fold, respectively, albeit to levels that were still about tenfold below those seen for wild-type VPS4A. Although the incomplete rescue by the $VPS4A_{K173Q/L64A,D}$ double mutants probably reflects a complex combination of competing effects, this experiment indicates that secondary Leu 64 mutations decrease the severity of class E compartment formation and thereby release endogenous ESCRT machinery to function in virus budding.

VPS4 enzymes are composed of three different functional elements: the N-terminal MIT domain, which binds the ESCRT-III lattice, a central AAA ATPase cassette, which mediates oligomerization and hydrolyses ATP, and a β-domain, which is inserted into the small domain of the ATPase cassette and is required for LIP5/Vta1 cofactor binding[27]. VPS4 enzymes assemble into dodecamers composed of two stacked hexameric rings, with MIT domains positioned above a large central chamber created within one of the two asymmetric rings (Fig. 4c; Z. Yu, G. Gonciarz, W. I. Sundquist, C. P. Hill and G. J. Jensen, personal communication). The ESCRT-III lattice is not yet fully characterized, but biochemical and crystallographic analyses[16,28–30] have shown that ESCRT-III protein oligomerization and membrane binding are autoinhibited by terminal elements that include the C-terminal helices described here. Removal of these elements promotes CHMP3 dimerization in solution and permits CHMP3 to be crystallized in protein 'strips' that contain two distinct dimer interfaces and create a continuous basic membrane-binding surface[28]. ESCRT-III membrane association may therefore serve to expose (and concentrate) the C-terminal CHMP1–3 helices. This oligomeric array of exposed binding sites would then promote the recruitment and assembly of dodecameric VPS4 enzymes by means of their MIT domains. The MIT domains, in turn, seem to reside at the ends of flexible tethers[27,28] and could therefore 'feed' ESCRT-III subunits into the open central chamber of the assembled enzyme.

METHODS SUMMARY

Biosensor binding experiments. Biosensor binding experiments were performed as described[11,26] with purified VPS4B MIT protein binding to immobilized GST–CHMP1B proteins (Fig. 1a) or C-terminal CHMP peptides binding to immobilized GST–VPS4 MIT proteins (Fig. 1b). All binding isotherms were fitted to simple 1:1 binding models.

Structural studies by NMR. Solution structures of VPS4A/VPS4B MIT–CHMP complexes were determined with NMR spectroscopic analyses of two sets of mixed isotope-labelled samples (all in fast exchange): $^{13}C,^{15}N$-VPS4A/VPS4B MIT plus excess $^{12}C,^{14}N$-$CHMP1A_{180–196}$/$CHMP2B_{195–213}$ and excess $^{12}C,^{14}N$-VPS4A/VPS4B MIT plus $^{13}C,^{15}N$-$CHMP1A_{180–196}$/$CHMP2B_{195–213}$. Bound structures were determined for the $^{13}C,^{15}N$-labelled component in each sample, and then docked to create binary complexes by using data from half-filtered nuclear Overhauser enhancement spectroscopy.

GFP–CPS sorting in yeast cells. *vps4Δ* yeast cells (SEY6210 genetic background) harbouring pRS415*MET25+GFP–CPS* (GFP–CPS expression) and either pRS416+*VPS4* or pRS416+$vps4_{L64D}$ (Vps4 expression) or an empty control plasmid were imaged by confocal microscopy and scored for GFP fluorescence that was primarily at the limiting vacuolar membrane or primarily within the vacuolar lumen. Yeast class E compartments and vacuolar membranes were revealed by staining with FM4-64.

GFP–VPS4 localization in mammalian cells. COS7 cells grown on coverslips were transfected with wild-type and mutant enhanced green fluorescent protein (EGFP)–VPS4A expression vectors[26]. Cells were fixed 18 h after transfection, imaged by confocal fluorescence microscopy, and scored as having either cytoplasmic or punctate GFP–VPS4 distributions.

HIV-1 vector release and infectivity. 293T cells were transfected with EGFP–VPS4 expression constructs together with an HIV-1 vector system (pCMVΔR8.2 plus pWPTS-nlsLacZ plus phCMV-VSVG). Cells and supernatants containing virions were harvested separately 48 h after transfection. Cellular protein expression levels and virion release were analysed by western blotting, and vector titres were determined by transducing HeLa-M cells.

Full protocols. Detailed methods describing sequence alignments, protein expression and purification, binding studies, NMR structure determinations, cell biology and virology experiments are provided in Supplementary Methods.

Received 2 June; accepted 15 August 2007.

1. Hurley, J. H. & Emr, S. D. The ESCRT complexes: structure and mechanism of a membrane-trafficking network. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 277–298 (2006).
2. Williams, R. L. & Urbe, S. The emerging shape of the ESCRT machinery. *Nature Rev. Mol. Cell Biol.* **8,** 355–368 (2007).
3. Morita, E. & Sundquist, W. I. Retrovirus budding. *Annu. Rev. Cell Dev. Biol.* **20,** 395–425 (2004).
4. Carlton, J. G. & Martin-Serrano, J. Parallels between cytokinesis and retroviral budding—a role for the ESCRT machinery. *Science* **316,** 1908–1912 (2007).
5. Babst, M., Sato, T. K., Banta, L. M. & Emr, S. D. Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p. *EMBO J.* **16,** 1820–1831 (1997).
6. Babst, M., Wendland, B., Estepa, E. J. & Emr, S. D. The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J.* **17,** 2982–2993 (1998).
7. Babst, M., Katzmann, D., Estepa-Sabal, E., Meerloo, T. & Emr, S. Escrt-III. An endosome-associated heterooligomeric protein complex required for mvb sorting. *Dev. Cell* **3,** 271–282 (2002).
8. Phillips, S. A., Barr, V. A., Haft, D. H., Taylor, S. I. & Haft, C. R. Identification and characterization of SNX15, a novel sorting nexin involved in protein trafficking. *J. Biol. Chem.* **276,** 5074–5084 (2001).
9. Ciccarelli, F. D. *et al.* The identification of a conserved domain in both spartin and spastin, mutated in hereditary spastic paraplegia. *Genomics* **81,** 437–441 (2003).
10. Tsang, H. T. *et al.* A systematic analysis of human CHMP protein interactions: Additional MIT domain-containing proteins bind to multiple components of the human ESCRT III complex. *Genomics* **88,** 333–346 (2006).
11. Scott, A. *et al.* Structure and ESCRT-III protein interactions of the MIT domain of human VPS4A. *Proc. Natl Acad. Sci. USA* **102,** 13813–13818 (2005).
12. Reid, E. *et al.* The hereditary spastic paraplegia protein spastin interacts with the ESCRT-III complex-associated endosomal protein CHMP1B. *Hum. Mol. Genet.* **14,** 19–38 (2005).
13. McCullough, J. *et al.* Activation of the endosome-associated ubiquitin isopeptidase AMSH by STAM, a component of the multivesicular body-sorting machinery. *Curr. Biol.* **16,** 160–165 (2006).
14. Agromayor, M. & Martin-Serrano, J. Interaction of AMSH with ESCRT-III and deubiquitination of endosomal cargo. *J. Biol. Chem.* **281,** 23083–23091 (2006).
15. Ma, Y. M. *et al.* Targeting of AMSH to endosomes is required for epidermal growth factor receptor degradation. *J. Biol. Chem.* **282,** 9805–9812 (2007).
16. Zamborlini, A. *et al.* Release of autoinhibition converts ESCRT-III components into potent inhibitors of HIV-1 budding. *Proc. Natl Acad. Sci. USA* **103,** 19140–19145 (2006).
17. Howard, T. L., Stauffer, D. R., Degnin, C. R. & Hollenberg, S. M. CHMP1 functions as a member of a newly defined family of vesicle trafficking proteins. *J. Cell Sci.* **114,** 2395–2404 (2001).
18. von Schwedler, U. K. *et al.* The protein network of HIV budding. *Cell* **114,** 701–713 (2003).
19. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6,** 519–529 (2005).
20. Takasu, H. *et al.* Structural characterization of the MIT domain from human Vps4b. *Biochem. Biophys. Res. Commun.* **334,** 460–465 (2005).
21. Bishop, N. & Woodman, P. TSG101/mammalian VPS23 and mammalian VPS28 interact directly and are recruited to VPS4-induced endosomes. *J. Biol. Chem.* **276,** 11735–11742 (2001).

22. Fujita, H. *et al.* A dominant negative form of the AAA ATPase SKD1/VPS4 impairs membrane trafficking out of endosomal/lysosomal compartments: class E vps phenotype in mammalian cells. *J. Cell Sci.* **116,** 401–414 (2003).
23. Bowers, K. *et al.* Protein–protein interactions of ESCRT complexes in the yeast *Saccharomyces cerevisiae*. *Traffic* **5,** 194–210 (2004).
24. Ward, D. M. *et al.* The role of LIP5 and CHMP5 in multivesicular body formation and HIV-1 budding in mammalian cells. *J. Biol. Chem.* **280,** 10548–10555 (2005).
25. Shiflett, S. L. *et al.* Characterization of Vta1p, a class E Vps protein in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279,** 10982–10990 (2004).
26. Garrus, J. E. *et al.* Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107,** 55–65 (2001).
27. Scott, A. *et al.* Structural and mechanistic studies of VPS4 proteins. *EMBO J.* **24,** 3658–3669 (2005).
28. Muziol, T. *et al.* Structural basis for budding by the ESCRT-III factor CHMP3. *Dev. Cell* **10,** 821–830 (2006).
29. Lin, Y., Kimpler, L. A., Naismith, T. V., Lauer, J. M. & Hanson, P. I. Interaction of the mammalian endosomal sorting complex required for transport (ESCRT) III protein hSnf7-1 with itself, membranes, and the AAA+ ATPase SKD1. *J. Biol. Chem.* **280,** 12799–12809 (2005).
30. Shim, S., Kimpler, L. A. & Hanson, P. I. Structure/function analyses of four core ESCRT-III reveals common regulatory role for extreme C-terminal domain. *Traffic* **8,** 1068–1079 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Rich, D. Myszka and S. Endicott for support; D. Winge for amino acid analysis; J. Shaw for reagents and expertise; and S. Alam for NMR expertise. W.I.S. received funding from the NIH.

Author Information Atomic coordinates for VPS4A MIT–CHMP1A$_{180–196}$, free VPS4B MIT and VPS4B MIT–CHMP2B$_{195–213}$ are deposited in the Protein Data Bank under accession numbers 2jq9, 2jqh and 2jqk, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.I.S. (wes@biochem.utah.edu).

LETTERS

Effective RNAi-mediated gene silencing without interruption of the endogenous microRNA pathway

Matthias John[1], Rainer Constien[1], Akin Akinc[2], Michael Goldberg[3], Young-Ah Moon[5], Martina Spranger[6], Philipp Hadwiger[1], Jürgen Soutschek[1], Hans-Peter Vornlocher[1], Muthiah Manoharan[2], Markus Stoffel[6], Robert Langer[3,4], Daniel G. Anderson[4], Jay D. Horton[5], Victor Koteliansky[2] & David Bumcrot[2]

Systemic administration of synthetic small interfering RNAs (siRNAs) effectively silences hepatocyte gene expression in rodents and primates[1–3]. Whether or not *in vivo* gene silencing by synthetic siRNA can disrupt the endogenous microRNA (miRNA) pathway remains to be addressed. Here we show that effective target-gene silencing in the mouse and hamster liver can be achieved by systemic administration of synthetic siRNA without any demonstrable effect on miRNA levels or activity. Indeed, siRNA targeting two hepatocyte-specific genes (apolipoprotein B and factor VII) that achieved efficient (~80%) silencing of messenger RNA transcripts and a third irrelevant siRNA control were administered to mice without significant changes in the levels of three hepatocyte-expressed miRNAs (miR-122, miR-16 and let-7a) or an effect on miRNA activity. Moreover, multiple administrations of an siRNA targeting the hepatocyte-expressed gene *Scap* in hamsters achieved long-term mRNA silencing without significant changes in miR-122 levels. This study advances the use of siRNAs as safe and effective tools to silence gene transcripts in animal studies, and supports the continued advancement of RNA interference therapeutics using synthetic siRNA.

Recently, it was reported that adeno-associated viral (AAV)-expressed short hairpin RNA (shRNA), when administered to mice, can result in profound toxicity, presumably by saturation of the cellular miRNA pathway[4]. AAV expression of shRNA seemed to elicit toxicological effects by interference of cellular pathways for miRNA biogenesis, including transport of nuclear miRNA precursors to the cytoplasm by the nuclear karyphorin exportin-5. shRNA-mediated toxicological effects were evident by marked reductions in cytoplasmic levels of mature miRNA, and occurred in a manner independent of both 'on-target' silencing and shRNA sequence. The findings of ref. 4 can be explained by viral vector expression of shRNA at high levels, because lower-level shRNA expression seems to be tolerated[4,5]. However, the implications for cellular delivery of synthetic siRNA that acts downstream of miRNA biogenesis have remained undetermined.

We interrogated the biochemical and functional status of the miRNA pathway after *in vivo* administration of synthetic siRNA formulated in liposomal nanoparticles such as those recently described[2,6]. Formulated, synthetic siRNA targeting two hepatocyte-expressed gene transcripts (si-Apob targeting apolipoprotein B (*Apob*) and si-FVII targeting factor VII (*F7*)) and an irrelevant gene (si-Luc targeting luciferase) were administered by a single intravenous bolus injection at different dose levels in groups of three mice. At both 2 and 30 days, silencing of on-target gene transcripts in liver homogenates was measured by the branched DNA assay[2], and levels of endogenous miRNA (miR-122, miR-16 and let-7a) were assessed by both northern blot and nuclease protection assays.

Administration of high-dose (5 mg kg^{-1}) si-Apob and si-FVII resulted in marked silencing of the hepatocyte-expressed genes to 22 ± 3% and 17 ± 2% relative to placebo-treated mice at the 2-day time point, respectively (Supplementary Fig. 1). The siRNA effects were dose-dependent, with intermediate silencing effects observed in the animals treated with a low dose (2 mg kg^{-1}) at the same time point. Consistent with previous findings[1,2], the RNA interference (RNAi)-mediated silencing effect was selective, with no measurable effect on *Apob* mRNA levels with si-FVII or si-Luc treatment. Furthermore, there was no significant effect on *F7* mRNA levels in mice treated with si-Apob or si-Luc siRNA. At the 30-day time point, *Apob* and *F7* mRNA returned to near normal levels. Treatment with siRNA was not associated with any observed toxicities over the 30-day time period.

To determine whether either ~80% silencing of two different hepatocyte-expressed genes or hepatocyte delivery of three distinct siRNAs results in dysregulation of miRNA biogenesis, levels of the liver-specific miRNA, miR-122, and the broadly expressed miRNAs miR-16 and let-7a were measured in liver tissue samples at both time points for all treatment groups at all dose levels. As shown in Fig. 1, there was no significant reduction in miRNA levels for individual-siRNA-treated animals, as assessed by northern blot assay. To quantify more accurately miRNA levels, a nuclease protection assay was used to measure simultaneously each miRNA together with the small nuclear RNA *U6*. Inclusion of the *U6* control allowed for normalization of the total amount of RNA in each assay. As summarized in Supplementary Table 1, no significant differences were measured in miRNA levels in siRNA-treated animals. Expression levels ranged from 80% to 110% of the levels measured in PBS-treated animals, and were not associated with any trend related to the siRNA dose level or degree of gene silencing. Certainly, no reduction of miRNA levels to match the >80% level previously reported in ref. 4 was observed with administration of synthetic siRNA.

The northern blot and nuclease protection results established that synthetic siRNAs do not inhibit the synthesis or processing of cellular miRNAs. We extended these findings to investigate whether systemically administered siRNAs interfere with endogenous miRNA function. Although *in vivo* data on miRNA function are limited to only a few published reports[7–9], it has been shown that inhibiting mouse miR-122 with antagomirs significantly increases the mRNA levels of a number of liver-expressed genes including

[1]Alnylam Europe AG, Fritz-Hornschuch-Str. 9, 95326 Kulmbach, Germany. [2]Alnylam Pharmaceuticals Inc., 300 Third Street, Cambridge, Massachusetts 02142, USA. [3]Department of Chemistry, and [4]Center for Cancer Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetss 02139, USA. [5]Department of Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9046, USA. [6]Institute of Molecular Systems Biology, Swiss Federal Institute of Technology ETH Zürich, HPT E73, CH-8093 Zürich, Switzerland.

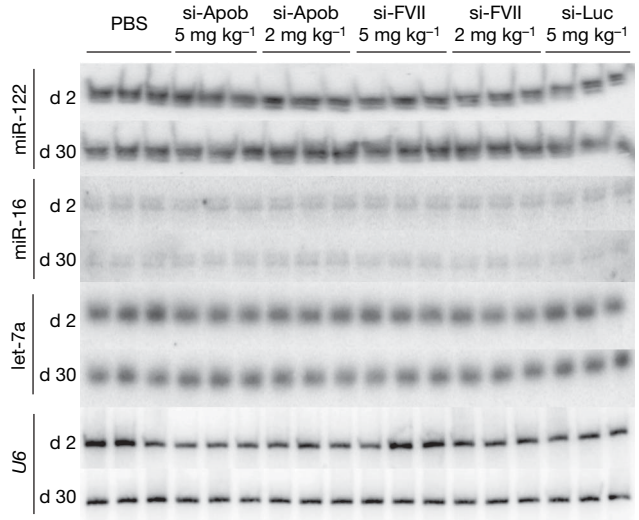


Figure 1 | Northern blot analysis of mature miRNAs reveals no detectable differences among siRNA treatment groups. Northern blots of total RNA from mouse livers isolated two days (d 2) and 30 days (d 30) after intravenous injections of PBS, si-Apob (5 mg kg^{-1} and 2 mg kg^{-1}), si-FVII (5 mg kg^{-1} and 2 mg kg^{-1}) or si-Luc (5 mg kg^{-1}) hybridized to radiolabelled probes specific for miR-122, miR-16 and let-7a (one animal per lane). Hybridization to a radiolabelled probe specific for *U6* was used to confirm equal loading of RNA.

Aldoa, *Hfe2*, *Tmed3*, *Lass6*, *Slc35a4*, *Tmem50b* and *Gpx7* (ref. 9, see Supplementary Fig. 2). Accordingly, liver mRNA levels for these seven putative miR-122 targets were quantified using branched DNA assays (normalized to *Gapdh*) in mice treated with si-Apob, si-FVII and si-Luc (two days after injection), and compared to PBS-treated animals. Moreover, protein levels of total Aldolase and another miR-122 target, Iqgap1 (ref. 9), were assayed by western analysis of liver lysates. Consistent with the lack of effect on mature miR-122 levels, there was no increase in the mRNA or protein levels of any of these genes that could be correlated with the dose of siRNA administered or the degree of gene silencing observed (Fig. 2 and Supplementary Fig. 3).

These studies in mice demonstrate that acute gene silencing of ~80% by single administration of siRNA is not associated with any changes in biosynthesis of miR-122, miR-16 or let-7a, nor with any increase in miR-122 target-gene expression. To explore whether

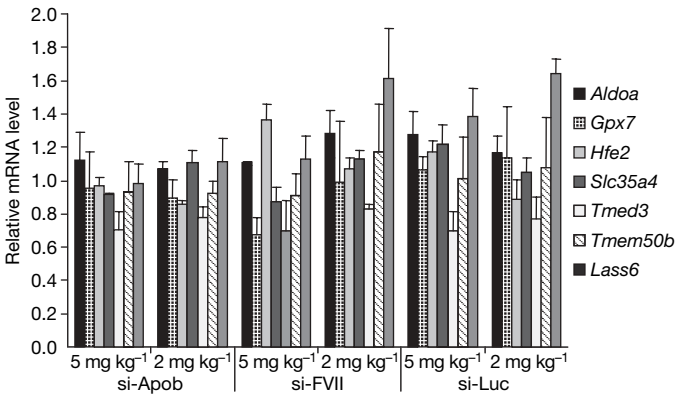


Figure 2 | Liver mRNA levels of miR-122 targets are unaffected by siRNA treatment. Mouse liver *Aldoa*, *Gpx7*, *Hfe2*, *Slc35a4*, *Tmed3*, *Tmem50b* and *Lass6* mRNA levels normalized to *Gapdh* mRNA measured 2 days after intravenous injections of PBS or different doses (5 mg kg^{-1} and 2 mg kg^{-1}) of formulated siRNA targeting *Apob* (si-Apob), *F7* (si-FVII) or luciferase (si-Luc) ($n = 3$ per group). Data are shown as group mean ± s.d. normalized to the PBS group for each gene.

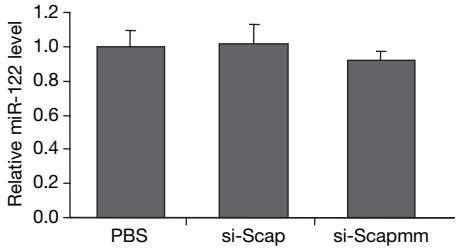


Figure 3 | Long-term silencing of *Scap* in hamster liver by repeat administration of siRNA does not affect miR-122 levels. Hamster liver miR-122 levels measured by nuclease protection (normalized to *U6* RNA levels) 7 days after three, weekly intravenous injections of PBS or 2.5 mg kg^{-1} formulated siRNA targeting *Scap* (si-Scap) or a mismatch control siRNA (si-Scapmm). Data are shown as mean ± s.d.

prolonged gene silencing by multiple doses of siRNA would alter miRNA levels, and to extend our studies to another species, we targeted the hepatocyte-expressed gene SREBP cleavage-activating protein (*Scap*) in hamsters. To silence *Scap*, hamsters were given three, weekly injections of a formulated siRNA targeting *Scap* (si-Scap) or a mismatch control siRNA (si-Scapmm) at a dose of 2.5 mg kg^{-1}. Repeated treatment with si-Scap resulted in significant (~75%) silencing of *Scap* mRNA, as well as a very significant reduction in Scap protein on day 21 relative to saline-treated animals (Supplementary Fig. 4). As shown in Fig. 3, silencing of *Scap* for three weeks by multiple doses of si-Scap had no effect on levels of miR-122, nor did repeat-dosing of the mismatch control, si-Scapmm.

Our findings demonstrate that robust gene silencing, both acute and longer-term, can be achieved in the liver without any detectable alteration of cellular miRNA biogenesis or function. Certainly, additional studies are warranted such as those with longer 'dicer substrate' siRNAs (>21 nucleotides in length) that act upstream of 21-nucleotide siRNA and could interfere with miRNA biogenesis[10]. Meanwhile, effective gene silencing with siRNA, in which RNA interference is mediated downstream of miRNA biogenesis, can be dissociated from toxicities that may be intrinsic to DNA-expressed shRNA.

METHODS SUMMARY

Antagomirs and siRNAs were synthesized as described[1,9]. Formulation of siRNAs into liposomes was performed by extrusion. Antagomirs and formulated siRNAs were administered to mice and hamsters by intravenous injection. Expression levels of miRNAs were determined by northern blotting and nuclease protection assay of total RNA isolated from livers of treated animals. Levels of miR-122 target mRNAs were measured in liver lysates using the QuantiGene assay (Genospectra). Western blot analysis was used to measure Scap protein levels in hamster liver lysates and miR-122 target (Aldoa, Iqgap1) protein levels in mouse liver lysates.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 November 2006; accepted 16 August 2007.
Published online 26 September 2007.

1. Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432**, 173–178 (2004).
2. Zimmermann, T. S. *et al.* RNAi-mediated gene silencing in non-human primates. *Nature* **441**, 111–114 (2006).
3. Song, E. *et al.* RNA interference targeting Fas protects mice from fulminant hepatitis. *Nature Med.* **9**, 347–351 (2003).
4. Grimm, D. *et al.* Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**, 537–541 (2006).
5. Narvaiza, I. *et al.* Effect of adenovirus-mediated RNA interference on endogenous microRNAs in a mouse model of multidrug resistance protein 2 gene silencing. *J. Virol.* **80**, 12236–12247 (2006).
6. Akinc, A. *et al.* A combinatorial library of lipid-like materials for delivery of Ray therapeutics. *Nature Biotechnol.* (submitted).
7. Esau, C. *et al.* miR-122 regulation of lipid metabolism revealed by *in vivo* antisense targeting. *Cell Metab.* **3**, 87–98 (2006).

8. Krutzfeldt, J., Poy, M. N. & Stoffel, M. Strategies to determine the biological function of microRNAs. *Nature Genet.* **38** (Suppl), S14–S19 (2006).
9. Krutzfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
10. Kim, D. H. *et al.* Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nature Biotechnol.* **23**, 222–226 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Wetzel, M. Jayaraman, K. N. Jayaprakash, K. G. Rajeev, L. Nechev and G. Wang for assistance in chemistry; R. Dorkin for liposome preparation; M. Duckman for assistance with the figures; and J. Maraganore for critical advice on the manuscript. M. Spranger is supported by CC-SPMD/ETH. M.G., R.L. and D.G.A. are supported by an NIH grant.

Author Contributions Experimental work was performed by M.J., R.C., A.A., M.G., Y.-A.M. and M. Spranger. Experimental design and critical advice was provided by M.J., R.C., A.A., P.H., J.S., H.-P.V., M.M., M. Stoffel, R.L., D.G.A., J.D.H., V.K. and D.B. The paper was written by D.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.B. (dbumcrot@alnylam.com).

METHODS

Preparation of siRNA. siRNAs were chemically synthesized as described[11], and formulated into liposomes by extrusion[6]. The sequence of si-Apob has been described[12]. Sequences of the other siRNAs used in these studies are as follows: si-FVII sense, 5′-GGA**UCA**U**CUC**AAG**UCUU**A**C**TT-3′; si-FVII antisense, 5′-G**U**AAGA**CUU**GAGA**U**GA**UCC**TT-3′; si-Luc sense, 5′-cuuAcGcuGAGuAcu-ucGATT-3′; si-Luc antisense, 5′-UCGAAGuACUcAGCGuAAGTT-3′; si-Scap sense, 5′-GcuuAAuGGuucccuuGAuTT-3′; si-Scap antisense, 5′-AUcAAGG-GAACcAUuAAGCTT-3′; si-Scapmm sense, 5′-GcuuAAucGuAccGuAGAuTT-3′; si-Scapmm antisense, 5′-AUcuAcGGuACGAUuAAGCTT-3′. Lower-case letters denote 2′-*O*-Me-modified nucleotides; bold letters denote 2′-*F*-modified nucleotides. All siRNAs contain phosphorothioate linkages between the two thymidines (T) at the 3′ end of each strand. The sequences and synthesis of the antagomirs has been described[13].

***In vivo* animal experiments.** Animal studies were performed with the approval of the Institutional Animal Care and Use Committee (Alnylam) and the Institutional Animal Care and Research Advisory Committee (University of Texas Southwestern). Saline, siRNA formulations and antagomirs were administered to mice by tail-vein injection, and to hamsters by the jugular vein. For formulated siRNA, the dose (given in mg kg^{-1}) refers to the amount of siRNA in the formulation. Mice received a single dose of siRNA or saline, and were killed 2 and 30 days after injection by CO_2 inhalation. Hamsters received three doses of siRNA or saline (days 0, 7 and 14) and were killed on day 21 by isoflurane overdose. For the antagomir study, mice received three, daily doses, and were killed 24 h after the third dose. Livers were dissected and snap-frozen for gene expression, miRNA and western analysis.

mRNA measurements. To measure levels of mouse *Apob*, *F7*, *Aldoa*, *Hfe2*, *Tmed3*, *Lass6*, *Slc35a4*, *Tmem50b*, *Gpx7* and *Ppia* (encoding cyclophilin) mRNA, the QuantiGene assay (Genospectra) was performed essentially as previously described[11] using primer sets specific for each gene. Levels of hamster *Scap* mRNA, normalized to β-actin mRNA, were measured by quantitative PCR with reverse transcription (RT–PCR). For hamster studies, total liver mRNA was isolated using the RNeasy kit (Qiagen). Complementary DNA was transcribed (ABI cDNA synthesis kit; PE Biosystems) from RNA and used in SYBR Green (PE Life Sciences) RT–PCR on an ABI PRISM 7900HT System (Applied Biosystems). *Scap* versus β-actin normalization was performed by the threshold cycle difference (ΔCT) method. Primer sequences were: Scap-5′, 5′-GTACCTGCAGATGATGTCCAT-TG-3′; Scap-3′, 5′-CTGCCATCCCGGAAAGTG-3′; β-actin-5′, 5′-GGCTCCCA-GCACCATGAA-3′; β-actin-3′, 5′-GCCACCGATCCACACAGAGT-3′.

Immunoblots. Total protein extracts were prepared from frozen, pulverized mouse livers, separated by SDS–PAGE (12%) and transferred to nitrocellulose membranes (Schleicher & Schuell) by electroblotting. Proteins were detected using specific antibodies for total Aldolase (Cell Signaling, Inc.), Iqgap1 (BD Transduction Laboratories) and tubulin (BD Transduction Laboratories). Horseradish peroxidase-conjugated secondary antibody and chemiluminescent detection (Amersham) were used to visualize the proteins.

Membrane proteins were prepared from frozen hamster livers as described[14]. Equal amounts of protein (30 μg) were subjected to SDS–PAGE on 8% gels and transferred to Hybond ECL membrane (GE Healthcare Bio-Sciences Corp.). Immunoblot analyses were performed using polyclonal anti-hamster Scap antibody, and anti-mouse receptor-associated protein antibody as described[15]. Antibody-bound bands were detected using the SuperSignal CL-horseradish peroxidase substrate system (Pierce Biotechnology Inc.).

Northern blotting. Synthetic RNA probes complementary to mouse miR-122, miR-16 and let-7a were labelled at their 5′-end using polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (GE Healthcare). Total liver RNA was isolated by Trizol (Invitrogen) extraction, dissolved in formamide loading buffer, run on denaturing 12% acrylamide gels and transferred to Hybond-N membranes (GE Healthcare) by semi-dry blotting. After baking at 80 °C for 3 h, membranes were equilibrated in 90% Express-Hyb solution (Clontech) and hybridized for 1 h with the radio-labelled RNA probes. After repeated washing, membranes were exposed to a phosphorimager screen and analysed on a Typhoon 9200 instrument (GE Healthcare).

Nuclease protection assay and miRNA quantification. Synthetic DNA probes complementary to mouse or hamster miR-122, mouse miR-16, mouse let-7a, and mouse or hamster *U6* RNA were labelled at their 5′-end as described above. Probe sequences were: miR-122, 5′-AAACACCATTGTCACACTCCACAGC-TCTCTCTTCT-3′; miR-16, 5′-GCCAATATTTACGTGCTGCTAAGGCACCG-CTTTCTTCTTCTTC-3′; let-7a, 5′-ACTATACAACCTACTACCTCATCCC-ACAGTGTTCTTCTTCTTCTTC-3′; and U6, 5′-GTGTCATCCTTGCGCAGG-GCCATGCTAATTTCTTCTTCTTC-3′. Total liver RNA was simultaneously hybridized in solution to a miRNA-specific probe and the U6 probe. The hybridization conditions allowed detection of *U6* RNA and mature miRNA, not pre-miRNA. After treatment with S1 nuclease, samples were loaded on denaturing 10% acrylamide gels. Gels were exposed to a phosphorimager screen and analysed on a Typhoon 9200 instrument (GE Healthcare). Relative signal intensities of miRNA versus U6 were calculated for each sample.

Statistical analysis. Statistical significance was calculated using ANOVA followed by Tukey's post-hoc analysis.

11. Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432**, 173–178 (2004).
12. Zimmermann, T. S. *et al.* RNAi-mediated gene silencing in non-human primates. *Nature* **441**, 111–114 (2006).
13. Krutzfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
14. Engelking, L. J. *et al.* Overexpression of Insig-1 in the livers of transgenic mice inhibits SREBP processing and reduces insulin-stimulated lipogenesis. *J. Clin. Invest.* **113**, 1168–1175 (2004).
15. Shimano, H. *et al.* Overproduction of cholesterol and fatty acids causes massive liver enlargement in transgenic mice expressing truncated SREBP-1a. *J. Clin. Invest.* **98**, 1575–1584 (1996).

CORRIGENDUM

doi:10.1038/nature06271

Southern Ocean sea-ice extent, productivity and iron flux over the past eight glacial cycles

E. W. Wolff, H. Fischer, F. Fundel, U. Ruth, B. Twarloh, G. C. Littot, R. Mulvaney, R. Röthlisberger, M. de Angelis, C. F. Boutron, M. Hansson, U. Jonsell, M. A. Hutterli, F. Lambert, P. Kaufmann, B. Stauffer, T. F. Stocker, J. P. Steffensen, M. Bigler, M. L. Siggaard-Andersen, R. Udisti, S. Becagli, E. Castellano, M. Severi, D. Wagenbach, C. Barbante, P. Gabrielli & V. Gaspari

Nature **440**, 491–496 (2006); doi:10.1038/nature04614.

In calculating the fluxes used in this paper, we made a mistake in the snow accumulation rate used. We calculated fluxes assuming that snow accumulation rates were expressed in metres water equivalent, but they were actually expressed in metres ice equivalent. The result of this error is that the fluxes shown in the original paper are uniformly 8.3% too high. The shape of the affected plots (the lower four panels in Figure 2, both panels in Figure 3, and the top and bottom panels in Figure 4) remains the same if the correct values are plotted, and the affected figures could be corrected by re-scaling the y-axes by 8.3%. Our conclusions are not affected. The flux data in the Supplementary Information of the original paper ('Supplementary Data 2') have now been corrected.

naturejobs

JOBS OF THE WEEK

In the 1999 comedy film *Office Space*, three guys get sick of working in a cubicle farm for an idiotic boss and a measly salary. They decide to rebel, and in one scene they abscond with the office printer — which never seems to work and causes no end of frustration. They carry the hapless machine out to a field and take turns hitting it with a baseball bat. To them, the printer represents all that is wrong with the office experience. Many who have worked in an office can empathize, including scientists who do their share of office work and have to deal with bureaucracy. What if there were a better way?

This week's Feature (see page 750) looks at the pluses and minuses of doing astronomy research from the comfort of your home. Some US astronomers are now making a living as freelancers, receiving grants directly, without involving a university. There are advantages: no commuting, the freedom to make your own schedule, fewer overhead costs than at a university, and no university bureaucracy.

But anyone who has done any sort of freelancing knows that there are also drawbacks to being completely on your own every day from nine until five. The daily rat race, although gruelling, also provides human interaction, structure and a bit of discipline. Working at home can get lonely — or not lonely enough, if you have children in the house craving attention. Freelance astronomers — or anyone else working for themselves — might benefit from the experiences of freelance writers. So here are some tips that I have accrued over the years.

Create a schedule and stick to it. Work every day for a specific number of hours — this is key to remaining productive and efficient. Make sure you get exercise; it is all too easy to remain inside all day. Network —you need to stay in touch with the community to keep abreast of job opportunities. And, perhaps most importantly, be aware of your financial situation. Freelancing requires dedication and hard work to keep multiple funding sources going so that your income remains steady and sufficient. Freedom from the daily commute and the office has its rewards. It has caveats as well.

Gene Russo, acting editor of *Naturejobs*

CONTACTS
Acting Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

FREEDOM OF THE SKIES

W. M. KECK OBSERVATORY

For astronomers looking for maximum independence, working from home is an option. But it has its challenges. **Genevive Bjorn** reports.

When astronomer Barbara Whitney moved to Madison, Wisconsin, to be with her partner, she had difficulty finding an astronomy job. She decided to switch research fields and pursue atmospheric physics. But working conditions were tough and the pay was low. Frustrated, she left science altogether and went to work for a small-business development centre.

But after a few years missing the fun of research, she started writing astronomy grant proposals. She won all of them. Her success was tempered, however, because she lacked the time and expertise necessary to deal with the complicated paperwork and regulations — so she could not actually receive the money. She decided to join the Space Science Institute (SSI), a non-profit organization in Boulder, Colorado, that manages research grants. Ten years on, Whitney has rebuilt a successful astronomy career.

A small but significant subset of US astronomers are foregoing the security of a university job for the freedom of working on their own. The SSI and a similar organization called the Planetary Science Institute (PSI) in Tucson, Arizona, say their ranks are growing. Anita Krishnamurthi, chair of the American Astronomical Society's employment committee, says there is increased interest in freelancing from postdocs and mid-career researchers — so much so that the topic is now regularly covered at career counselling sessions during the society's meetings. Rejecting the typical academic track, the freelancers conduct research from the comfort of their homes. The benefits are many, but so are the risks. Funding is a constant pursuit, and access to an actual observatory and graduate students are even harder to come by.

Astronomers who become self-employed are generally looking for flexibility, says Krishnamurthi.

M. HOLAHAN/METRO

Two-body problem: Mike Wolff (top) and Barbara Whitney found solutions.

She estimates that there are about 200 self-employed astronomers, most of whom left university positions but wanted to keep doing their science. The trend began in the early 1970s when some bright young planetary scientists were seeking independent research without university obligations. They set up the PSI in 1972. Since then, several other non-profit institutes have formed, giving astronomers in need of flexibility more choices for support — and more are being set up.

"The Internet has been key," says Krishnamurthi. "Now that all the journals are online and everything is web-based, you don't need on-campus library access any more." The nature of the work also lends itself to independence. Astronomers tend to form far-flung research partnerships, so working with collaborators remotely is nothing new. Most astronomers do not require sophisticated equipment or lab space on a daily basis — just the occasional use of a telescope. Most of the work is done on a personal computer.

The independent career path is similar to non-tenure-track research jobs at universities. The transition into the field is relatively easy for scientists already working on 'soft money' — that is they are entirely funded by grant money as opposed to the dedicated salary received at, for example, a university by a tenure-track faculty member — who want more flexibility and freedom.

How to get started

So far, the independent option is available only in the United States. Self-employed astronomers fund their own salaries and research through grants and contracts offered by organizations such as NASA or the National Science Foundation. These funding agencies place tight restrictions on accounting and management. Individual scientists cannot receive funds directly; rather, an institution such as the SSI of the PSI must

receive and manage grants. The scientists must spend their grant money in approved ways, and the institution must provide regular, detailed financial reports to the funding agency. And anyone receiving money from the US government has to provide proof of compliance with federal acts enforced by the Equal Employment Opportunity Commission, such as the Equal Pay and Civil Rights Acts.

There is an upside, however. Because organizations such as the PSI and the SSI do not have large campuses or bureaucracies to support, the overheads cost as little as half those found at universities. Most provide employment benefits, such as health insurance and retirement plans, for an additional fee.

Outside the United States, there are no organizations like the PSI and the SSI, mainly because funding mechanisms are not in place. Government funds go directly to government agencies, such as the CNES space agency in France, or to public–private partnerships such as Germany's Max Planck Society. Astronomers have the chance to work as independents on soft money only if they come to the United States under a suitable work visa. "Many astronomers work part of the year in other countries, but government funding does not cross borders," says PSI director Mark Sykes.

"Places like the SSI really allow you to do what you want to do from wherever," says Mike Wolff. "If you bring your own money, you have a job." He describes himself as an offsite employee who pays his own way. For Wolff and his family, it solved the oft-encountered two-body problem. His wife had a great job offer in Augusta, Georgia, but he had no prospects there. Unwilling to live apart from her and their children, or to give up astronomy, Wolff obtained his own funding and joined the SSI, where he is now co-director of research and a senior research scientist.

The independent track can help parents deal with family concerns. Heidi Hammel, of Ridgefield, Connecticut, juggles the demands of three children with a full-time astronomy career. As an independent, she can deal with crises in both areas as they happen. "If a proposal deadline is rapidly approaching, I'll work on that," she explains. "But if the school calls and tells me my kid has head lice, I drop everything and go."

That's one of many ways the working conditions are better than in traditional settings. Recovering from an absence, in terms of pay and promotion, can be difficult at a university. Committee meetings, interruptions and excess bureaucracy reduce productivity and add stress.

"Because I am not very self-promoting, I tended to accept lower-paying, lower-status jobs," Whitney says. "By going independent, I was able to catch up in pay with a position that I deserve." She believes that in the university system, it would have been hard to get back to the level she was on before her break.

Working from home also brings challenges, regardless of the field. Feelings of loneliness and isolation can set in after the glow of uninterrupted time has worn off. And it can be difficult to set boundaries with family members, especially young children. Wolff says his three school-age children frequently interrupt him, making his work less efficient and his hours irregular, especially during the summer. He often gets things done while travelling. A regular programme of exercise, talking with colleagues and a good home-office set-up are key to staying productive, he says.

A. COOPER

The sky's the limit: being a freelance astronomer can offer a lot of flexibility.

Follow the money

But perhaps the biggest potential downside, Krishnamurthi notes, is that you are completely responsible for your own salary. The funding process for independent astronomers is a source of both freedom and stress. Fluctuations in the funding climate create job insecurity and keep some people from taking the risk of striking out on their own. Most independent astronomers devote at least 10% of their time every day to writing grant and contract proposals. "We worry about funding constantly, but somehow manage to hang on and develop successful careers," says Hammel. "But the reality at universities is that you don't have job security either, unless you are tenured."

Independent astronomers can choose their own salaries and give themselves regular pay rises when they have sufficient grant money — a move that must be documented and in-line with industry standards. They need to bring in less money than typical faculty members as the overheads are lower, and they don't generally have to support postdocs, graduate students and a summer salary. "This is a valuable and satisfying career path as long as people are able to handle funding uncertainty and go into it knowing what is involved," says Krishnamurthi. Those who take the risk and become self-employed tend not go back.

Astronomers wanting to pursue independent research find more success if they have done one or more postdocs, made a name for themselves and had some experience writing grant proposals. "When you've been around for a while, with more experience and knowledge of how the field works, you are in a position to make a better choice than walking into it straight out of graduate school," says Krishnamurthi.

Tom McCord, former research scientist at both the PSI and the SSI, now director of the Bear Fight research and meeting centre in Washington, looks at it another way. "Mentoring with old farts like me is a way graduate students and postdocs can get into the game," he says. It is possible to strike out on your own as a postdoc, he adds. But hustling for funding, studious grant writing and an appreciation of the freedom and flexibility are essential for any real hope of succeeding. ■

Genevive Bjorn is a freelance writer based in Maui, Hawaii.

Tom McCord helps to mentor postgraduates.

Web links

Eureka Scientific
www.eurekasci.com
Planetary Science Institute
www.psi.edu
Space Science Institute
www.spacescience.org
Southwest Research Institute
www.swri.org

NATURE|Vol 449|11 October 2007

MOVERS

Elena Shevchenko, scientist, & Elena Rozhkova, assistant scientist, Center for Nanoscale Materials, Argonne National Laboratory, Illinois

Elena Shevchenko (far left)
2005–07: Staff scientist, the Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California

Elena Rozhkova
2005–07: Research professional associate, University of Chicago, Illinois

Despite divergent scientific training, Elena Shevchenko and Elena Rozhkova have much in common besides their given name. Both grew up in Russia, did postdocs in the United States and moved to Chicago to find new opportunities. Each is now eager to tap into the other's expertise as they join Argonne National Laboratory's new NanoBio Interfaces group at the Center for Nanoscale Materials (CNM) to jointly develop novel materials.

Shevchenko is the nanotech expert. She studied synthesis of magnetic nanoparticles for her PhD at the University of Hamburg in Germany before turning to self-assembly of nanoparticles at both the IBM Thomas J. Watson Research Center and Columbia University in New York. She pioneered the use of nanometre-scale forces and charges to create structures with novel optical and electrical properties from dissimilar inorganic materials such as semiconductors and gold. Her success with superlattice structures prompted a move to the Molecular Foundry at Lawrence Berkeley National Laboratory in California. Now she wants to integrate biological components into her inorganic research.

"We'd like to do design multicomponent materials using both inorganic substrates as well as viruses or proteins to see what properties can be extracted from such materials," says Shevchenko.

Rozhkova brings the biological perspective to the group. A bioorganic chemist, she left Russia after receiving a fellowship from the Japan Society for the Promotion of Science to study biological electron-transfer systems. She then moved to Princeton University in New Jersey to continue her studies of biological reactions. Then, at the University of Chicago, she began work on nanocomposites based on biologically inspired processes. "It has been a challenge throughout my career to decide whether I wanted to work in bioorganic synthesis or chemical biology," says Rozhkova. "But I feel those combined skills will serve us well as we develop new materials."

Although Argonne will focus their research efforts on energy-related problems, these nanobio materials are likely to find multiple uses. Eric Isaacs, director of the CNM, is eager to explore the potential of these materials not only in solar energy conversion, but as 'smart' composites that can deliver an inorganic particle to destroy a specific cancer cell. Given the wide-ranging, multidisciplinary goals of this new programme, Isaacs says Argonne is fortunate to recruit two scientists with such complementary talents. ■

Virginia Gewin

MENTORS & PROTÉGÉS

Awards for South Africa's best

Ubuntu is an African concept of mutual support under a communal umbrella. It's a principle embodied by the winners of *Nature*'s 2007 South Africa Mentoring Awards.

Barry Fabian, who won the lifetime achievement award, and mid-career recipient Charles de Koning are professors at the University of the Witwatersrand, Johannesburg. Both returned to South Africa after training abroad. Nominators praised them not only for their scientific acumen but also for their ability to pass it on without condescension, and for their equitable treatment of all, regardless of race, sex or background.

Fabian was inspired by his own mentor, Margaret Kalk, who graded his essay from jail after being arrested on an anti-apartheid protest. Now an emeritus professor of developmental biology, he says that academics in South Africa have a responsibility not just to train young scientists, but to create an informed scientific culture within government, the civil services, the media and the public. Jane den Hollander, pro-vice-chancellor of Curtin University of Technology in Western Australia, remembers struggling to adapt as an undergraduate from a small mining town. With Fabian's encouragement, she went on to do her MSc with him. Later she was offered a position outside academia. "Everyone but Barry thought I was making a great error," Hollander says. Fabian asked her about her skills, strengths and what she wanted in life: indirect guidance that helped her make a satisfactory decision.

De Koning has had 32 MSc and PhD students, who have helped him produce 61 of his 75 published papers. Perhaps the greatest testament to his efforts is the number of his students who stay in science. Many undergraduates are pressured to leave with a BSc, then work in industry to support their extended families. He's sympathetic, but emphasizes the rewards of postgraduate education.

One student had tended cattle and worked in the mines before coming to Witwatersrand. With constant encouragement from de Koning, he eventually did a PhD in Germany and a postdoc in Atlanta, Georgia, before returning to South Africa.

De Koning's commitment is summed up by Edwin Mmutlane, a research chemist at the Council for Scientific and Industrial Research in Pretoria. "On a continent characterized by a dearth of skills and scientific role models, he is doing a phenomenal job as a relatively young person who could have otherwise opted for the alluring prospects of the developed world," says Mmutlane. ■

Paul Smaglik

POSTDOC JOURNAL

Perseverance pays off

I've just heard that the final bit of research from my PhD has been accepted for publication. I'm obviously pleased, but I feel more like sighing in relief than punching the air. It's taken me two-and-a-half years, and two major rewrites, to get to this point — and there's only so many times I can read and recraft the same sentences before I start to get jaded.

From a professional perspective, however, I still have to hold my breath, because publication of my work is just the beginning. My data and ideas are now out in the public domain for other scientists to see and scrutinize. For my paper to be considered successful, that's exactly what I need them to do: read, consider and use what I've written to inform their own research. (Ideally, they'll also agree with my conclusions, but I'm certainly not averse to a good argument.)

Unfortunately, if past experience is anything to go by, I'll have to wait at least a couple of years before I know if I've made an impact or not. It seems funny that, six years since starting my PhD research, I'm still waiting to see how it will turn out. Perhaps the true key to academic survival is not enthusiasm or intellectual brilliance — although both help — but patience and stubborn perseverance. ■

Chris Rowan is a postdoctoral student in the geology department at the University of Johannesburg, South Africa.

NATURE|Vol 449|11 October 2007

BYOB FAQ

It's all you've ever wanted.

Terry Bisson

Where do you get your BOBs?
From you! Each and every BOB is unique, custom designed to order for each client. Your personal BOB, neurally mapped to your specs from an approved and tested blank, will be unique.

Are they really volunteers?
We wouldn't have it any other way. Our programme begins and ends with free choice, yours and his. Our blanks are Asian and African males, aged 28–36, who have freely chosen to have their personalities erased and remapped (not just overwritten) in order to have a chance at a new life in the United States or Europe.

How are they selected?
With special care. We accept only healthy mature male blanks, HIV- and STD-clean, cosmetically and medically reconditioned before being neurally reconfigured to make a satisfactory boyfriend, life partner or husband, if you so desire.

How do you know what kind of guy I'm looking for?
You tell us. Simply ENTER your own needs and preferences into our proprietary matching database. That's all there is to it! Unlike older programs, *Build Your Own Boyfriend*™ lets you choose exactly the qualities you want in a life-partner. And then delivers it.

Will he remember his previous life?
Your BOB comes with a full generic set of memories that is chemically dimmed, giving him a feeling of completeness without the specificity of individual recollection. You and your BOB will immediately begin making your own memories. That's what relationships are all about! Although many of our blanks come from penal or military points of origin (POOs), they have been completely erased, not just overwritten, before reconfiguration.

What if he doesn't like me?
Unlikely, as your BOB is configured to like the same things you do — which includes yourself! And in the unlikely event that you are not satisfied (in every way) with your delivered BOB, you are free to return him at any time in the first six months after reception, with only a nominal restocking fee and no questions asked. This happens in only a small percentage of cases. These BOBs are returned to inventory to be rewritten and reassigned. They have no memory of their reception. You have no responsibility for a returned blank.

Can I choose race or ethnicity?
Sorry. BYOB operates under strict non-discrimination laws. We guarantee only that your BOB will be healthy and pleasant looking with no disfigurements. Most are Asian or African, as EU restrictions prohibit European blanks at present.

JACEY

What if I don't want a long-term relationship?
Then our service is not for you. To adopt a metaphor from the stock market, we are not day traders, nor do we short-sell. We're here for the career women willing to make an investment in a long-term partner. Have we mentioned that our BOBs are covered by our Comprehensive Health Insurance policy (CHI)? You can and should look forward to a long and satisfying relationship.

Men only? What's that about?
We do service select gay men, but our service is primarily for women seeking a long-term companion or partner. Current international sex-traffic ordinances prevent our acceptance or reconfiguration of female blanks.

Will I have to teach him English?
Not at all. Your BOB comes with mature language capabilities, which are independent of memory. He may be teaching you! As our Syntax module is based on the Webster–Chomsky proprietary syntax map. He will, however, be unable to read or write. Many clients regard this as a plus.

What about citizenship?
Each BOB is awarded conditional citizenship six months after reception. It is among the many things he will be thankful to you for! And although BOBs cannot vote or own property, they have most of the rights of unconditional citizens. As an added attraction, our Canadian or US clients are granted an extra one-fifth vote. A similar concession is currently being negotiated with the government of the United Kingdom.

Will he long for his old life?
No. He will remember only that it was unpleasant and will be neurally incapable of remembering any specific incidents or people. His new life with you will be all that is of interest to him.

Will he know that he is a BOB?
Only if you tell him. He will know only that he has a past personality that he is disinclined (and indeed unable) to access. Many women find pleasure telling their BOBs that they have been especially designed to suit them. Many BOBs find comfort in the knowledge that they are 'special' in this way. But again, it's your call.

What if I grow tired of him?
Why would you? Remember, your BOB wasn't just matched with you, he was made for you. But in the unlikely event that you wish to discard your BOB at any time after the initial BYOB warranty period, you can do so without legal prejudice by delivering him to Migration Control, as his citizenship is conditional and he is neurally mapped to go without resistance. Your responsibility then ends, and he is free to become blanked again, or be returned to his Point of Origin (POO).

What if I have further questions?
They can be submitted in confidence to our BYOB website at www.Bob.bio; or if you wish to speak to a live operator, 1 (919) 456 8999. Now, may we ask you a question?

What are you waiting for? Haven't you been lonely long enough? Share your life with a BOB who is designed to fit your lifestyle and unique personality. Send for our Profile Initiator today! ■

Terry Bisson is a Hugo-award winning science-fiction writer based in California. His novella *Dear Abbey* was a nominee for the British Science Fiction award in 2003.